NPS-OR-95-015

NAVAL POSTGRADUATE SCHOOL
Monterey, California

UNITED STATES NAVY
NAVAL POSTGRADUATE SCHOOL

SOVIET VISUAL PERCEPTION RESEARCH: APPLICATION TO TARGET ACQUISITION MODELING

by

Judith H. Lind

December 1995

Approved for public release; distribution is unlimited

Prepared for: U.S. Army Training and Doctrine Analysis Command
White Sands Missile Range, NM

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943-5000

Rear Admiral M.J. Evans
Superintendent

Richard Elster
Provost

This report has been prepared for and funded by the U.S. Army Training and Doctrine (TRADOC) Analysis Command, White Sands Missile Range, New Mexico. The report has been reviewed by Timothy Linn and Peter Shugart, TRADOC Analysis Command; and Ronald A. Erickson, ASI Systems International, Ridgecrest, California.

This report was prepared by:

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE December 1995	3. REPORT TYPE AND DATES COVERED Technical

4. TITLE AND SUBTITLE Soviet Visual Perception Research: Application to Target Acquisition Modeling	5. FUNDING NUMBERS
6. AUTHOR(S) Judith H. Lind	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943	8. PERFORMING ORGANIZATION REPORT NUMBER NPS-OR-95-015
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Training and Doctrine Analysis Command White Sands Missile Range, NM 88002-5502	10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Five Soviet books have been reviewed to ascertain how target acquisition was modeled in the former Soviet Union and to determine if information is sufficient to program a comprehensive model. Authors include V.D. Glezer and K.N. Dudkin of the Pavlov Institute of Physiology, St. Petersburg. Since the books (published between 1961 and 1985) were machine-translated from the Russian, some original concepts may have not been correctly interpreted. Still, they provide an excellent overview of 30 years of vision research at the Pavlov Institute and of Russian thought on vision and the brain.

The Soviet texts emphasize cognitive mechanisms of vision more than is common in U.S. military models. Mental models and the observer's mindset are considered very important. More emphasis is given to modeling recognition and identification (versus detection) than in the U.S. The result of this study is a sketchy and incomplete search and target acquisition model, unsuitable for programming at present. The reviewed books mostly provide information about vision in general, with emphasis on proposed neurophysiological and psychological processes that may explain experimental results. They obviously were not written with computer programs in mind. Extensive data collection would be required to quantify the Soviet vision concepts for use in a computer model.

14. SUBJECT TERMS Combat models; Detection; Dudkin, K.N.; Glezer, V.D.; Human factors; Human performance; Identification; Memory; Mental models; Modeling and simulation; Pavlov Institute; Recognition; Russian models; Search; Soviet models; Target acquisition; Vision models; Visual perception	15. NUMBER OF PAGES 70 16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

This page intentionally left blank.

Contents

Contents (Continued)

Contents (Continued)

This page intentionally left blank.

Executive Summary and Evaluation

Five Soviet books and conference proceedings have been reviewed to ascertain how vision and target acquisition were viewed and modeled in the former Soviet Union. The goal of this project has been to determine whether sufficient information is available in these five documents to prepare a comprehensive model that might be programmed and compared with U.S. and British models that predict human target acquisition performance. Emphasis is on what may be unique to the Soviet model of vision and target acquisition.

The primary authors whose work has been reviewed are V.D. Glezer and K.N. Dudkin, respected researchers at the Pavlov Institute of Physiology in St. Petersburg, considered one of the world's foremost institutions working in the mathematical theory of vision. The reviewed documents were published between 1961 and 1985. All had been translated from the Russian (some by machine, and none by vision experts), so there is some uncertainty whether the authors' original concepts have been correctly interpreted during the translation process.

The result of this study is a very sketchy and incomplete model, unsuitable for programming at present. The reviewed books mostly provide information about vision in general (they are quite good graduate-level college texts), with emphasis on proposed neurophysiological and psychological processes that may explain experimental results. They obviously were not written with computer programs in mind; very few programmable algorithms or equations are included. Extensive data collection would be required to quantify the Soviet vision concepts for use in a computer model.

Nonetheless, the concepts in these books are important and do provide an excellent overview of Soviet thought on vision and the brain. Thus a typical sequence of target acquisition events is used here to organize various model factors and components considered important by Glezer, Dudkin, and their associates (see *Conclusions and Tentative Top-Level Model* at the end). The various factors have been separated into the categories of detection, aimpoint, recognition, and identification depending on the process they affect (with several factors included under more than one category). The resulting list may serve as a framework for eventual development of a comprehensive model, if additional data can be located or generated.

In summary, the following points can be made.

1. The results of the Pavlov Institute's 30 years of vision research are impressive; U.S. vision researchers are encouraged to obtain these writings (and others that now may be available) for comparison with U.S. theories and research results.

2. The Soviets have made good use of Information Theory, Pattern Recognition Theory, Decision Theory, and similar concepts in developing their general vision models.

3. Considerably more emphasis is placed on cognitive mechanisms of vision, especially as these relate to target acquisition, than is observed in U.S. military models. Mental models and the observer's mindset appear to be considered as important as physiological processes of vision for detection and identification. More emphasis is given to modeling recognition and identification processes (versus detection) than in the U.S.

This page intentionally left blank.

Introduction

Background

The dissolution of the Soviet Union and opening of technical documents to the West provides us with the opportunity to learn the nature of Soviet concepts of the target acquisition process. That is, it is possible now to review translations of seminal books and papers written by vision and cognitive science experts which provide the theoretical and practical background for understanding how search, detection, and identification have been viewed in Eastern Europe. That is the goal of this report.

A total of five books and conference proceedings have been reviewed for this study. These are

1. V.D. Glezer, *Information and Vision*, USSR Academy of Sciences I.P Pavlov Institute of Physiology, 1961.

2. *Information Processing in Visual Systems: Proceedings of the 4th Symposium on Sensory System Physiology*, Leningrad, November, 1976.

3. V.D. Glezer and K.N. Dudkin, *Visual Identification and its Neurophysiological Mechanisms*, USSR Academy of Sciences I.P Pavlov Institute of Physiology, 1975.

4. K.N. Dudkin, *Visual Perception and Memory*, USSR Academy of Sciences I.P Pavlov Institute of Physiology, 1985.

5. V.D. Glezer, *Vision and Thought*, USSR Academy of Sciences I.P Pavlov Institute of Physiology, 1985.

A goal of this project has been to determine whether sufficient information is available in these five documents to prepare a comprehensive model that might be programmed and compared with U.S. and British models that predict human target acquisition performance. The reviewed documents were published between 1961 and 1985. All had been translated from the Russian (some by machine and generally not by vision experts), so there is some uncertainty whether the authors' original concepts have been correctly interpreted during the translation process.

Since these books have been reviewed with the specific goal of locating and documenting information that can be used to develop models of human target acquisition performance, less attention has been paid to the results of animal experiments that are not directly applicable to human vision. Experimental data and very general information that cannot be included in a practical model of the search and target acquisition process also are not discussed here.

The summaries provided here generally do not include information about the vision system and human vision processes that are widely known and are included in most U.S. models. Emphasis is on what may be unique to Soviet concepts of vision and target acquisition and on background information needed to understand their concepts.

Definition of Target Acquisition

For purposes of this study, military target acquisition is considered to be a complex process that includes several more-or-less distinct components, discrimination levels, or tasks. Although these components are not defined precisely in this manner in the Soviet literature, it is obvious from the reviewed documents that some similar process is considered in their vision modeling. It should be noted that the five acquisition components listed below generally are thought of as occurring serially; each depends on successful completion of the previous step. That is, the search process culminates in target detection, which is followed sequentially by aimpoint location, target recognition, and finally target identification. Target acquisition is a general term that includes all five items defined below. Definitions of these terms are taken from several U.S. military reports.[1, 2, 3]

1. **Search**. Active scanning of the field of regard; the goal is to locate a target of military interest. Search is a more-or-less continuous *process*; the other four acquisition tasks are better described as discrete *events*. Search ends (at least temporarily) when a object of interest is located.

2. **Detection/location**. Sensing that an object that is foreign to the background is in the sensor's field of view (FOV) and should be further examined to determine if it is a target of military interest. The object may have been visible before detection, but was not distinguishable enough from other objects to trigger the inspection decision. The observer now takes whatever action is necessary to inspect the object (e.g., rotate the sensor, change to narrow FOV).

3. **Aimpoint**. Selecting a portion of the observed scene as a potential target and a specific point at which to aim. The observer begins to move his weapon into attack position.

4. **Recognition**. Selecting a particular spot in the scene as the target on the basis of characteristics of its shape, coupled with categorizing a military target by class, such as a vehicle (or, if the class is taken to be more specific, a tank, truck, or personnel carrier). The level of detail for recognition depends on the operational situation and on prebriefing. The observer now begins attack mode, including possibly designation of the object.

5. **Identification**. Recognizing that the military target is in a specific subclass (e.g., tank) or is a specific member of a class (e.g., T-72 rather than a T-62). The subclasses are dependent on classes, the operational situation, and prebriefing. The observer continues preparation for attack and commits to weapon firing or release.

Target acquisition is a cognitive process that must be reported or otherwise indicated by the observer, if it is to be recorded. Performance levels can be judged in several ways, including (1) percent of targets acquired, (2) range at which targets are acquired, (3) cumulative percent of targets acquired as a function of range, (4) time required to acquire, (5) the number of objects falsely identified as targets, and (6) the accuracy of locating a target on a map.

[1] Naval Air Warfare Center Weapons Division. *Target Acquisition Models for Janus (A)*, by J.H. Lind, NAWCWPNS, China Lake, CA, Feb. 1995 (NAWCWPNS TM 7811).

[2] U.S. Army TRADOC Analysis Center. *Janus (T) Documentation*. TRAC, White Sands, NM, June 1986.

[3] Naval Weapons Center. *Review of Mathematical Models of Air-to-Ground Target Acquisition Using TV and FLIR Sensors*, by A.D. Stathacopoulos and others, General Research Corp. NWC, China Lake, CA, Jan. 1976 (NWC TP 5840).

Information and Vision

Introduction

Information and Vision was published in 1961, the work of Vadim Davydovich Glezer. Glezer has been the head of the Pavlov Institute of Physiology in St. Petersburg, considered one of the foremost groups working in the mathematical theory of vision.

The focus of this book is application of Information Theory methods to the understanding of image sensing by the human optic system. A single theoretical-practical approach to biological and technical systems of image transmission is sought, leading to general rules concerning the transmission of optical images and the understanding and modeling of the visual process. Glezer suggests that the results then can be applied to development of automatic pattern recognition systems.

Glezer notes that applicable experimental research is very sparse, and hypotheses must substitute for data in many instances. Thus he made no attempt to include color or stereoscopic vision in his 1961 models.

Quantizing Images in the Optic System

The amount of information that can be transmitted from the eye to the brain per unit time depends on visual contrast sensitivity, visual resolution, and temporal resolution (resolution over time). Although Glezer considers the luminosity of objects to be a continuous function, he sees visual perception as discrete. That is, images consist of a finite collection of discrete elements. The optic system quantizes constant images and transmits them to the brain as discrete samples. Thus the eye cannot discern changes that occur in very short time periods.

On the positive side, discretization of the visual image reduces the negative impact of intermittent interference (quantum fluctuation of the light source, dark noise of the retina, and optic system nerve channel noise). Still, it is necessary for the optic system to isolate useful light signals from noise.

Light Sensitivity Thresholds

Luminance values are given in apostilbs in Glezer's book. These have been converted to more familiar cd/m^2 units for this report (1 apostilb = 3.1416 cd/m^2). For reference, *photopic vision* (chromatic cones active) requires luminance values that range from just below the upper limit of visual tolerance (between 10^5 and 10^6 cd/m^2) to about 0.1 cd/m^2. *Mesopic vision* (both rods and cones active) occurs with luminances from about 0.1 to 10^{-4} cd/m^2. *Scotopic vision* (achromatic rods active) is experienced when luminance is below about 10^{-4} cd/m^2, down to the absolute threshold of visibility at between 10^{-6} and 10^{-7} cd/m^2.

Physiological optics usually separates light thresholds into three categories.

1. **Absolute threshold.** The absolute luminance threshold L_A is defined as the minimal detectable magnitude of luminosity on a "β-threshold" light spot against a black background under conditions of dark adaptation.

2. **Difference or incremental threshold.** This is defined as ΔL, the *difference* in luminosity between the light spot L_S and background L_B if the spot is on an illuminated (not black) background:

$$\Delta L = L_S - L_B .$$

3. **Differential threshold or contrast threshold**. This is defined as the *ratio* of the difference threshold ΔL to background luminance L_B. According to the Weber-Fechner Law, this ratio is constant at intermediate background luminosities, from approximately 0.3 to perhaps 300 cd/m^2 (1 to 1,000 apostilbs). That is, for this wide range of photopic luminosities, the difference threshold ΔL increases as the background luminance L_B increases:

$$\frac{L_S - L_B}{L_B} = \frac{\Delta L}{L_B} = \text{Constant} .$$

For lower values of background luminance in the low photopic, mesopic, and high scotopic visual ranges (3 x 10^{-1} to 3 x 10^{-4} cd/m^2), the differential threshold ratio rises rapidly (from about 0.2 to 1.0) as background luminance decreases.

In complete darkness, the observer sees weak luminescence or "dark noises" due to spontaneous impulses or fluctuations in the inherent light of the retina. For the rods, this luminescence is equivalent to luminous flux of about 1,000 quantum/second degrees^2. For the cones, the luminescence is 10^3 to 10^5 times greater. The level of dark noise differs widely for individual viewers. However, it is significant only at very low light levels, far below those normally accompanying target acquisition.

Retinal Receptive Fields

Light intensities are transmitted through the visual system as a finite number of discrete levels. The number of such levels is not fixed, nor is the threshold of perception. Instead, the illumination threshold I_T of the perception of a luminous spot depends on the size S of that spot:

$$I_T * S = \text{Constant} .$$[4]

This is referred to as Ricco's Law or the Law of Complete Summation. The latter name comes from indications that the actions of all light quanta that fall on a given section of the retina somehow are summed. A dimly-illuminated large spot can have the same effect as a brightly-illuminated small spot.

The size of the *zone of complete summation* depends on its location on the retina, background brightness, stimulus color, and presentation duration. Threshold stimulation depends only on the number of quanta of light absorbed by the cell, not on how the quanta are distributed within the summation zone.

The fovea is responsible for transmittal of most of the information that passes through the eye. In the center of the fovea and under dark adaptation conditions, the zone of complete summation has been found to be between 3 and 7 arc-min, depending on the color of the stimulus. Using white light, Glezer measured the zone to be 5.4 arc-min (standard deviation 0.71). At the edge of the fovea, the zone can be as much as twice this size. Incomplete summation occurs if stimulus size exceeds those values.

The retinal periphery serves primarily for detection of objects and determination of their direction of movement. In the visual periphery of the dark-adapted eye, the zone of complete summation has been reported by various researchers to be from 15 to 31 arc-min at various locations on the retina. Glezer notes that the size of the zone is different for different observers, and drops from about 30 arc-min at the fovea to 1 arc-deg at a distance of 12 degrees from the fovea. The summation capacity of the rods generally increases with distance from the center of the retina to the periphery. This explains the dark-

[4] The multiplication operator will be indicated by an asterisk (*) in all equations (standard in modeling and programming) to minimize possible confusion with the variable *x*.

adapted eye's ability to observe a dim signal peripherally but not foveally, and also the degradation of the signal's resolution due to the large area of luminance that is summed in the periphery.

Spatial summation of light in a light-adapted eye is insignificant or non-existent, when compared to that of the dark-adapted eye. The spatial summation of the rod apparatus changes minimally as light increases. For the cone apparatus under conditions of high illumination, the zone of complete summation approaches the angular size of a single cone, 0.4 arc-min. As illumination decreases, the size of the zone increases. In the region from about 3 x 10^{-2} to 30 cd/m^2 (mesopic vision) the following empirical relationship holds:

$$(\Delta L + L_B) * S * \sigma^a = \text{Constant}.$$

The value $(\Delta L + L_B) * S$ is proportional to the quantity of light that falls on a given section of the retina, not to exceed the size of the zone of completion summation σ. The parameter a is close to 1.

An area of the retina whose cones and rods all are tied to a specific ganglion cell or its fiber (and that thus respond as a functional unit) is referred to as a *receptive field.* Reactions of the entire receptive field are not equal to the simple sum of the reactions of its separate sections. The ganglion cell or *accumulating cell* that makes up a receptive field collects light from its associated rods and cones. In the visual periphery some ganglion cells occupy an area 350 microns in diameter, closely agreeing with the rod zone of complete summation of 1 arc-deg (in the retina, 1 arc-min is equal to 5 microns).

Thus Glezer concludes that the zones of complete summation for the rod apparatus are located in correspondence with the receptive fields associated with the branching of large ganglion cells, and has proposed that the convergent paths of neurons in the optic system are responsible for the zone of summation. Similar conclusions have been reached for foveal vision.

The receptive field appears to be integrated with the ganglion cell and accumulates the entire luminous flux that falls on its receptors. The zone of complete summation could be independently isolated from the background if the optic system of humans was a photometer that made absolute luminosity measurements. But the optic system is a photometer of comparison that allows only comparisons of intensities, not absolute magnitudes.

Research demonstrates that the zone of complete summation for the cone apparatus is not constant; it changes within a wide range as illumination changes, whereas rod fields change only insignificantly. A receptive field in the fovea can change approximately 40 times in size, ranging from 1 to 40 cones in a field at a time.

For foveal vision under mesopic light conditions (3 x 10^{-2} to 3 x 10^{-3} cd/m^2), when the stimulus spot lies within one receptive field, the visual threshold is constant. That is, it does not change as background luminance increases, for a given spot size. When the light stimulus extends beyond the borders of a single field, a gradual increase in threshold value is seen with increasing luminance.

Under photopic illumination conditions (above 0.03 cd/m^2), when the stimulus spot size is no greater than a single receptive field, the following relationship holds:

$$(\Delta L + L_B) * S \approx \sqrt{L_B}\ .$$

Under the same photopic illumination conditions, when the stimulus spot size is larger than a single receptive field, the relationship changes to

$$\Delta L + L_B \approx L_B\ .$$

The area of the receptive field σ is a function of the luminosity of the background, according to the following relationship:

$$\sigma \approx \frac{1}{\sqrt{L_B}}.$$

That is, the brighter the background, the smaller the receptive field (the more it contracts), decreasing from about 10 arc-min in the mesopic range to less than 1 arc min at 30 cd/m^2.

When the stimulus spot is approximately equal to the size of the receptive field and the background is between 0.1 and 20 cd/m^2 (low-light-level photopic vision), it is observed that

$$\frac{\Delta L}{L_B} \approx 1.$$

That is, the threshold contrast is constant and close to 1 in the luminance area where receptive field size varies with luminous intensity. This is independent of the size of the receptive field. In this range (0.1 to 20 cd/m^2), the number of luminance gradations in the human receptive field can be calculated.

If $\frac{\Delta L}{L_B} = 1$, then the difference threshold is

$\Delta L = L_B = 0.1 \text{ cd}/\text{m}^2$ for the first gradation, and

$\Delta L + L_B = 0.2 \text{ cd}/\text{m}^2$ for the next gradation, and so on.

Each subsequent difference threshold is twice the previous one. The m-degree threshold will correspond to a luminosity of $(0.1 * 2^{m-1})$ cd/m^2. Calculating m for the maximum light level in this range (20 cd/m^2), Glezer finds that the number of luminance gradations in the human receptive field is approximately 9 for low-light-level photopic vision.

Temporal Accumulation

The total number of quanta of light that fall on one accumulating cell (ganglion) depends not only on the luminous flux on the area, but also on the amount of time that the stimulus is located within the receptive field. Thus both spatial and *temporal summation* (or *temporal accumulation*) occur in the visual system. Temporal summation depends both on the area of the light stimulus and on its intensity The threshold effect is not observed if the time of stimulus presentation t is too short, that is, below a critical value t_{cr}, called the *critical duration*. The critical duration for a receptive field is of the order of 0.1 second at low light levels. At higher illuminations this value decreases. When t > t_{cr}, visual acuity is proportional to the logarithm of the object's illumination.

It has been observed that the perception of a signal depends both on the level of illumination I_T (stimulus size and luminance) and on the presentation time t according to the following equation:

$$I_T * t = \text{Constant}.$$

If the illumination level changes over the presentation time, the corresponding integral is used:

$$\int_0^t I_T dt = \text{Constant}.$$

This law is observed only when the signal presentation time does not exceed t_{cr}, the critical duration or summation time. The effect does not depend on how the quanta are distributed over time, but only on the value of the total number. Several authors have determined minimum values for the time summation process ranging from 0.0015 to 0.05 second.

The foveal cone field has about 1,000 times fewer sensors than the peripheral rod field, along with a much smaller area. However, the critical durations for these two kinds of fields are almost identical.

Related to temporal summation is the *time of inertia* t_I which depends on a dimensionless extinction function $A(t)$. This is the time the image or other visual perception is maintained in the optic system after it disappears from the field of view:

$$t_I = \int_0^\infty A(t)dt\,.$$

For the center of the fovea, t_I is about 0.012 to 0.2 second. For foveal vision in general, time of inertia is approximately 0.2 second. For peripheral vision, values ranging from 0.1 to 0.32 second have been measured.

The time of inertia has been shown to decrease with the intensity of the stimulus, ranging from 0.01 to 0.1 second in the eel; although the values may not hold for the human, the relationship appears to be true. Temporal summation times can change by 1.8 times (even as the receptive field's zone of complete summation can change in size by about 40 times).

The Statistical Character of Visual Thresholds

The value of the visual threshold is not constant, but rather fluctuates around some average value, based approximately on a Poisson distribution. When the light stimulus overlaps several receptive fields, the same “report” is sent to the brain via several channels (nerve fibers). Fluctuations in the values reported by the various channels are independent. When the same report is transmitted simultaneously along two channels, the relative scatter (noise) decreases by $\sqrt{2}$. This is observed during transfer from monocular vision to binocular vision. If the stimulus overlaps N fields, then a decrease in the threshold value by $\sqrt{N}$ is expected. However, experiments on humans have shown notable deviations from this expected value during summations for long periods of time and over large areas. Glezer notes that psychophysical experiments on humans are necessary for defining the statistical nature of the visual system, including temporal and illuminance summation.

In summary, in Glezer’s model, the human visual analysis system can be considered a discrete system characterized by (1) definite thresholds for discerning illumination, (2) receptive field size, and (3) the critical duration of a light stimulus.

Information Contained in Images

Glezer considers that, in the human eye, the retina transmits discrete images to the brain. The apparent continuity of image luminance and contrast is an illusion of vision. The quantization of images is determined by the functional organization of the neuronal system, not by the eye's resolution capacity or morphological structure.

As a simple model, an image can be considered a distribution of various magnitudes L of luminance on a surface. That is, the image is a function $L(x, y)$ where luminance at the independently-changing x and y locations can increase or decrease within given ranges. L can be considered the luminosity, brightness, and flow density in the conductors that stimulate a nerve cell. For a more complex model, an image can be considered a function of three variables $L(x, y, t)$, to include *time* for a mobile image. Other functions can be generated to include variables for a target's apparent size or color.

Returning to the simple model, *L(x, y)* may represent a continuous two-dimensional image that is discretized into a finite number of elements. A sinusoidal distribution *L(x)* can be used to describe the resolution of the optic device (human eye or electronic system) that is providing the continuous image:

$$L(x) = \frac{L_{Max}}{2}[\sin(2 * \pi * \nu * x) + 1].$$

The parameter x represents the distance along the x-axis of a specific portion of the image (that is, along its length; similar conclusions can be drawn with respect to the object's height along the y-axis). If a is the frequency or wavelength of the sinusoid, $\nu = 1/a$ is equal to the number of black (or white) strokes that subtend the image's length along x. The maximum number of strokes that can be resolved by the optical device along the image's length is called the *resolution capacity, W*.

The contrast resolution of an optical device can be defined by

$$C_R = \frac{L_{Max} - L_{Min}}{L_{Max}}.$$

The contrast of the image is zero if $\nu > W$ (that is, if one sinusoidal revolution is greater than the length of the image). The image displays contrast across its length whenever $\nu \leq W$. The image generated along the x-axis by the optical system (within its specified limits of resolution) next can be defined as a series of points (or strokes) that follow each other in intervals of $1 / (2 * W)$, with the contrast of each point either the same as or different from that of the preceding one.

The luminosity of the image *L(x)* is measured at each of the points that define the image's length. Linear interpolation between the points results in "smoothing" of the contrast levels at each point, as the image's actual continuous contrast variations are replaced with discrete levels of luminosity *L(x)* at the points. The number of discrete, independent readings of which a given sensor is capable is referred to as the sensor's *degrees of latitude* and is determined to be $2 * W * T$, where the signal's duration is T and sensor's bandwidth is W.

The image luminosity *L(x)* can take on only a discrete number of values, a function of the sensor's sensitivity. For most non-color sensor images, the image is discretized into strokes or lines, and the luminosity takes on only two values: white and black.

The fovea contains only cone cells. Under night vision conditions, each cone's receptive field is about 5 arc-deg in diameter. The fovea then contains about 200 such receptive fields, and the total angular subtense of the fovea is on the order of 90 arc-deg. The visual periphery contains primarily rod cells. Under night conditions, each of the 3,000 rod receptive field subtends about 1 arc-deg. The visual periphery is responsible for both detection and observation of weakly illuminated objects under night conditions.

With increasing illumination, the accumulative cone-type receptive cells assume the primary role in vision. Cone receptive fields shrink in size until, at about 30 cd/m^2, each receptive field is the size of a cone cell. In the periphery, increased brightness inhibits the rod cells and leads to increased activity of the cone cells located there, which respond as do the foveal cone cells. The greatest quantity of cone receptive fields in the periphery is located within an angular distance of 50 to 60 arc-deg from the fovea, in which area these fields total approximately 90.

Thus, under good daylight illumination conditions, the retina consists of about 800,000 receptive fields — approximately equal to the number of nerve fibers in the human optic nerve. The number of receptive fields in an area of the retina is approximately proportional to the square of that area's visual acuity. The resizing and reconstruction of the visual fields, which leads to changes in visual acuity, can be interpreted as transformations of the optic channel capacity.

Modeling an Image

A model describing the discrete structure of a transmitted image can be quite complex. However, a relatively simple model will be adequate if it describes a system that contains no less information than that actually received by the optical system. Such a model of the image should contain the following elements.

- Image elements that are identical in size, grouped according to their sizes based on the size of the smallest detail that can be discerned by the eye.
- The number of discernible luminance gradations.
- The number of discrete luminance changes that occur at less than visual flicker fusion frequency.

Information Capacity

It is often desirable to represent a multi-shaded image as a collection of binary numbers. To achieve this goal, an image can be described in tabular form. Each cell of the table corresponds to an element of the image. The value in the cell represents the brightness of that element. In the simplest case, the brightness is either *on* or *off.* For more complex systems, various gradations of brightness can be described using some numerical process, but usually still based on a binary system.

Using Information Theory terminology, the representations for the various discernible gradations of brightness are called *symbols* (e.g., white, light gray, dark gray, black), and the collection of symbols is referred to as an *alphabet*. The *information capacity* C of the information carrier is given as

$$C = \log_2 N = \log_2 m^n = n * \log_2 m .$$

The quantity N is the number of possible conditions (modeled brightness gradations, in this case), which is equal to m^n if the system consists of n accumulative cells or receptive fields, each with an identical number m of possible brightness gradations.

Information capacity thus grows rapidly (linearly) with the number of accumulative cells n involved and more slowly (logarithmically) with the possible levels of brightness m. The information capacity of a four-cell image that consists of two brightness levels ($C = 4 * \log_2 2 = 4$) is the same as a one-cell image that consists of 16 brightness levels ($C = 1 * \log_2 16 = 4$). But modeling a single accumulator that can possess 16 separate conditions is much more difficult than modeling four times that number of cells, each of which has only four possible conditions.

Information Needed for Image Transmittal

The information capacity level provides the defined (theoretical) capacity of the information carrier. The actual amount of information transmitted may be considerably less. The term *entropy* (H) is used to express the minimum information capacity of a channel that will be sufficient for transmittal of a given image.

If the image consists of an alphabet of two independent symbols (brightness levels) and the probability of the occurrence of one is p, the probability of the other is $(1 - p)$. Then the minimum information capacity of the channel needed to transmit one symbol is

$$H = -(p * \log_2 p) - (1 - p) * \log_2 (1 - p) .$$

If p equals either 1 or 0, the indeterminacy is absent and the entropy is 0. That is, if only one event is possible, any report about the image will contain no information.

If the alphabet consists of m possible symbols with probabilities $p_1, p_2, \ldots p_m$, then the minimum information attached to one symbol is

$$H = -(p_1 * \log_2 p_1) - (p_2 * \log_2 p_2) - \ldots -(p_m * \log_2 m) .$$

This means that, during the reporting of an event the probability of which is p_i, the growth in information is equal to ($-\log p_i$).

If the report (image) consists of N symbols or possible brightness levels, then the minimum information capacity required for its representation is equal to $N * H$. This is true because calculation of H results in an averaging of information transmission across all possible events. However, the term is exact only for large values of N.

When all symbols are equally probable, the maximum value of H, $H_{max} = 1$, is obtained. If the probability of one of two symbols is 1/16 and that of another is 15/16, the value of H drops to 0.278.

The degree of *redundancy* of information is expressed as

$$R = 1 - \frac{H}{H_{max}} .$$

Redundancy is equal to zero only when $H = H_{max}$. For example, when $H = 0.278$, $R = (1 - 0.287) = 0.722$. This information is useful in evaluating the effectiveness of information compression or image compression techniques.

To reduce redundancy, a shaded image can be replaced by an *outline* of the same object, with areas marked where there is a transition from black to white and from white to black. Since the frequency of transitions can be relatively low, the number of outline elements can be correspondingly low. The quantity of binary numbers needed to represent the object can be significantly smaller than for the original shaded image.

The above discussion applies only when information elements are statistically independent. However, for most images, elements demonstrate strong statistical ties, especially between neighboring elements. The brightness levels of adjacent elements usually are the same or differ by only one shade. Knowing the luminosity of one element, one can predict the value of an adjacent one with high probability of correctness. Studies of television and motion picture images indicate that only a small fraction of the total number of eight- and ten-shade images change intensity by more than one shade from frame to frame.

When the receiver of an image is the human visual system, the information received from images increases significantly. The number of discernible brightness levels depends on the size of the observed object. This number is greater for large homogeneous sections and smaller for borders, outlines, and small details.

Statistical Coding of Images

Principles of Statistical Coding

In the language of Information Theory, quantizing of information in the visual system is the same as the process of determining the *alphabet* of *symbols* used for an image. Information is created by the source of the reports (discretized images of continuous objects in the outside world). The information must be coded in order to transmit it from the source to the higher regions of the human visual analyzer. Conversion of the distribution of light on the retina to the distribution of the photo-receptors is the first stage of coding.

If coding did not occur, nearly 1 million samples of image brightness (based on the number of fibers in the optic nerve) would pass through the optic system every 0.1 second. The entire capacity of the brain would be expended in a matter of minutes. However, Glezer noted in 1961 that the coding mechanisms had not yet been defined. Several hypotheses had been proposed (mostly for information compression and reduction of redundancy), but no firm conclusions had been reached.

Decorrelation with Forecasting in the Visual System

Statistical redundancy of images is defined by the probable interactions, that is, by the correlations, between the elements. Codifying of images can begin with the elimination or decrease in these interactions, referred to as *decorrelation* of images. Two decorrelation techniques are used. The *forecasting method* predicts whether one element is followed by an identical or a different element, and ignores everything except changes. The *enlargement method* groups similar elements of the image, then codes them as a unit.

Decorrelation of images leads to a change of their statistical qualities. In their initial state, luminosity gradations can be considered equally probable when averaging a large number of elements over an image. The value of decorrelation lies in the fact that it is simpler to decrease or eliminate statistical redundancy if the redundancy is defined by inhomogeneity of the "non-linear" distribution of probabilities. After decorrelation, the unequal distribution of information in the image is quite noticeable. More information is carried by new values that are encountered infrequently. Decorrelation will permit selection and coding of those high-information components for transmission, versus attempts to code and transmit all or a uniform distribution of elements.

Decorrelation is tied to *inductive inhibition* of the retinal cells which results in reconfiguration of the receptive fields. As has been noted, light sensitivity is greater in the center of the visual field than in the periphery. The greater the illumination, the stronger the inductive inhibition of the periphery, thus concentrating the stimulus on the fovea and decreasing the effective area of the stimulus.

The impulse signals formed by the ganglion cells control information and also carry information concerning brightness changes to the higher regions of the visual analyzer. Reconfiguration of the foveal receptive fields requires less than 0.1 second, if illumination change is not too great. When brightness levels are sufficient (10 to 50 cd/m^2), a receptive field can be constricted to a single cone cell.

Decorrelation also is related to *receptor adaptation* and eye movements. Reacting only to changes is a general quality of the nervous system. If the intensity of stimulation is not varied, the frequency of nerve impulses reporting that stimulation gradually decrease and impulses sometimes cease. This process is referred to as adaptation of receptors. In the visual system, it is possible that specific receptive fields adapt over time also. Immobile images projected onto the human retina soon cease to be discriminable. Humans could not see static objects without the fine oscillatory movements that occur naturally in the eye. Several types of fine eye movements have been described: (1) tremor, (2) drift, and (3) saccadic eye movements or jumps. These movements have vertical, horizontal, and even rotary components.

Tremor has an amplitude of about 1 arc-min or less, equal to the angular size of a retinal photoreceptor. Although the tremor frequency bandwidth normally is 30 to 80 cycles per second, a continuous spectrum of frequencies from 1 to (at the extreme) 150 Hz has been reported.

Drift is a slow, smooth movement of the eye, shifting between 1 and 5 arc-min at a speed of about 1 arc-min per second. Drift does not seem to play a role in visual perception, but rather is associated with the instability of the eye movement system. The point of fixation for an image varies over an area with a diameter of about 10 arc-min (sometimes 20 arc-min). The eye returns to the point of fixation with a jump, after drift.

Saccadic eye movements are executed at high speed, usually completed in about 0.025 second. The amplitude can vary from 1 to 20 arc-min and occasionally rises to 50 arc-min. Saccades are not of equal size over the retina, but rather reach a maximum at the center of the fovea. Time between saccades ranges from 0.03 to 5 seconds, with 0.3 second being the most common duration for the fixations that occur between saccadic movements. Angular velocity can reach 400 arc-deg per second. Nothing can be observed during a saccade. However, only about 3% of observation time is required for these jumps, with the vision fixed for the remaining 97% (ignoring fine movements). Saccadic movements occur primarily during observation of large objects, as the eye jumps from one fixation point to the next. These movements are coordinated and synchronized with great accuracy across both eyes.

If an observer must track an object visually and that object moves at a constant speed not exceeding 30 arc-deg per second, saccadic movements are used first to direct the eye to the object. Then the eye tracks the object with an angular speed corresponding to the object's angular speed. Saccadic movements are used to correct for tracking errors. If an object is moving faster than 30 arc-deg per second, the eye is unable to keep up with its motion and saccadic movements are used to make up for the accumulation of tracking errors. If the object's motion is a random process, changes in tracking movement speeds are discrete, and occur in less than 0.1 second (apparently related to the critical duration during which an image must be present on the retina for observation).

When images are exposed on the retina for *less* than the critical duration (0.1 second), a stabilized (non-moving) image on the retina gives the sharpest resolution. The larger the amplitude of motion, the worse the image resolution. However, when the image is exposed for *longer* than the critical duration, the opposite is observed. A stabilized image then has poor resolution, and image sharpness increases with increasing amplitude of image movement. It appears that one of the functions of natural eye movements is de-adaptation and the resumption of *on* and *off* effects.

Borders between light and dark areas are especially important due to these eye movement effects. During tremor, the border moves relative to the receptors and de-adaptation occurs in that region. Between borders, the brightness level does not change and transmissions from these areas cease.

Possibly significant to detection are the neural system induction processes called *simultaneous contrast* and *subsequent contrast*. The former describes the situation where a gray object on a black background will appear lighter than the same object on a white background. The latter refers to the accentuation of various luminosities that are changed from one to another over time. These phenomena play a primary role in modifying receptive field sizes during observation of quickly changing images or those presented for short durations (less than the duration of saccades).

Perceptible luminosity is related to these two inductive processes, and is defined as

$$L_P = L_c + [f_1 * (L_{n_1})^{m_1}] + [f_2 * (L_{n_2})^{m_2}] - \{f_3 * [(L_{n_2})^g + (L_{n_1})^g]\} .$$

This describes the perceptible luminosity L_P of a test object with luminance L_{n_1} in the presence of an inductive object with luminance L_{n_2}. In the absence of inductive processes, $L_P = L_c$. That is, the perceptible luminosity is equal to the luminous contrast L_C. Otherwise, the perceptible contrast is increased due to neural induction (second and third components), and decreased due to reduction of the inhibition effect due to interactions of the inductive objects (this latter factor is very small). The constants f, g, and m depend on the placement, relative sizes, and forms of the inductive elements.

Visual interpolation of an image on the retina occurs, as is evidenced by the fact that we do not see the "blind spot" on the retina where the fibers of the optic nerve and blood vessels are located. Occupying nearly 6 arc-deg, and located only about 15 arc-deg from the fovea, the area has no photoreceptors. Yet, under ordinary circumstances the eye interpolates over this area so the observer is oblivious to its presence. Interpolation also can occur over large areas and even over the entire field of vision (for example, when observing a cloudless sky).

Coding of Images

An image with several emphasized outlines appears to have greater contrast. This is true even when outlines occur only in one direction, but even more so with outlines in both directions. Image resolution in general improves. Glezer considers that the outline carries the primary information concerning an object. Sections within and between outlines carry little information.

Increase in outline size or contrast allows the viewer to separate much of the information from noise. Knowledge of the outline often is sufficient for recognition of the visual form. A figure without substantial outlines located on a background with approximately constant luminosity will appear pale, weak, and fluctuating. If an outline is drawn around the figure, it immediately becomes well defined, with greater contrast and resolution. The primary function of receptive fields seems to be the detection and recognition of outlines, Glezer suggested in 1961.

Most of the information contained in an outline is related to those points where the outline changes direction sharply. Eye fixations not only follow an image's outlines, but also are more frequent on image sections where direction changes markedly and in areas of fine detail. Fixations are rarely observed where weak details are located.

Information is coded in the retinal receptive fields as a series of discrete impulses with equal amplitude, transmitted through the ganglion cells. The frequency of the impulses is proportional to the logarithm of the illumination of the retina. Perceived illumination thus is similarly proportional. However, one can postulate that the frequency does not matter, but rather the number of impulses in the sample is the important factor. The actual receptive field stimulant is not the illumination level but the amount of energy. The number of impulses is proportional to the logarithm of the light energy that falls within the effective areas of the receptive field.

Other kinds of information besides illumination changes are coded in the receptive field. These include the distribution of the intervals between impulses.

Glezer considers that there are two types of receptive fields. One type is at the retinal level, where receptive fields separate the signal from noise via accumulation, and decorrelates images using forecasting and elemental codification. The second type executes the more complex functions of detection and codification of simple configurations, the elements of visual forms. These receptive fields are tied to higher regions of the visual analyzer. This second type of field usually codifies a shape as an entire simple configuration, not element by element. Decorrelation is carried out using the enlargement process.

An image that occupies the field of vision can be described by a set of more or less complex forms. The entire set of shapes recognized by an individual form his or her alphabet. This complete alphabet divides itself into a series of partial alphabets that are stored in complex relationships of "hierarchical coordinates." The simpler the alphabet (that is, each symbol carries less information), the quicker it can develop. A complete set of forms, the alphabet of the visual analyzer, is not inherent. Rather, it is acquired during life experiences.

In summary, Glezer's model assumes that two important operations are carried out at the retinal level: accumulation in the receptive fields and decorrelation of images (or at least partial elimination of the statistical bonds between elements). Transmittal to the cortex of all elements of an image would be ineffective. The task of the visual analyzer is to extract information from the image, and to transmit those items that meet the analyzer's inherent criteria. Information derived from the images also must be stored. Effective coding is needed for both transmission and storage.

Throughput Capacity of the Optic System

The throughput of information in the optic system in a given period of time is limited by the activity of noise. A system that transmits error-free reports at a speed determined by the channel's throughput capacity is called an *ideal information transmission system* or an *ideal communication system*. The optic system is quite close to meeting the requirements for such a communications system.

Throughput Capacity of a Noisy Channel

The *speed of information transmission* R_u is the average value of the growth of information for one report, taking into account information unreliability:

$$R_u = H(x) - H_y(x).$$

The parameter $H(x)$ is the entropy of the distribution of probability values at the input of x in the system, and $H_y(x)$ is the unreliability, a measure of the indeterminacy of what was transmitted.

The *throughput capacity* C of a noisy channel is calculated as the maximum value of information transmission speed:

$$C = \max[H(x) - H_y(x)].$$

This maximum value corresponds to the best possible agreement of what was sent and what was transmitted along the channel. If the channel has a transmission frequency bandwidth of W at average transmitter strength, and if the desired signal is P_c, and if interference consists of white noise (Gaussian distribution) with an average strength of P_n, then

$$C = W * \log_2\left(\frac{P_c + P_n}{P_n}\right).$$

If H is the entropy of the report source (expressed in bits of information) and C is the throughput capacity of the communications channel, then no coding method exists that will guarantee a lower unreliability than $(H - C)$.

Optic Channel Capacity

The capacity of the optic channel is the amount of information that can be transmitted during the critical duration. If n is taken to be the number of receptive fields or independent nerve fibers, capacity can be calculated as

$$C = \frac{n}{2} * \log_2\left(1 + \frac{P_c}{P_n}\right).$$

If the number of independent optic nerve fibers is taken as 830,000, the maximum convergence of flashes is 55 Hz, and each fiber is assumed to transmit 1 bit of information for each flash, the total optic channel capacity is calculated to be

$$C \approx (830{,}000 * 55) \approx 45{,}000{,}000 \text{ bits/second.}$$

However, research indicates that the actual throughput capacity of the optic channel is on the order of one millionth of this quantity.

Visual Analyzer Throughput Capacity

The throughput capacity of the visual analyzer does not exceed a few dozen bits per second, the amount that can be perceived by the visual system as a whole. Although the retina is capable of sending tens of millions of bits of information per second, the higher levels of visual analysis do not receive information at this rate due to losses and information coding that occur at each step of neural transmission. The number of objects perceived simultaneously in a short period of time (less than the critical duration) is of the order of 7 ± 2. Information is stored in short-term memory without change for about 0.27 second, then disappears.

Research on identification of simple images with strong borders and good contrast indicates that at least 0.08 second is necessary for perception that an image is present. An object's general outline can be recognized in about 0.25 second. The object then can be identified if it is presented for a total of about 0.6 second. Throughput capacity for the images used was on the order of 50 to 70 bits per second.

Other research has resulted in throughput capacity values for various tasks. Reading speed (taking into account the statistics of language) is on the order of 30 to 40 bits per second. When adding two same-value numbers, throughput is 12 bits per second; subtraction of one number from another occurs at 3 bits per second.

With one-time presentation of a group of letters, the observer will perceive, on the average 4.5 letters, each of which contains 4.3 bits of information. Then

$$C = 4.5 * 4.3 / 0.27 = 72 \text{ bits / second .}$$

At low illumination levels the throughput capacity grows linearly with logarithmic increase of illumination. An illumination increase of 2 results in throughput capacity growth of approximately 10 bits per second. This probably is due to increases in visual channel capacity with increasing brightness. At higher levels of illumination, throughput capacity becomes constant.

Detection thresholds of the human visual system, at which the specified portion a of correct responses is observed, can be described by the following equation:

$$D_t = (p_1 - p_2)_a = k * \frac{p * q}{\sqrt{N}}.$$

The parameters p_1 and p_2 represent, respectively, the probability that a given point on an image is part of one of the bands (horizontal or vertical) that make up the test object image and the probability that the point is outside any band and simply represents noise. N is the number of points in the test object, $p = [p_1 + (p_2 / 2)]$, $q = (1 - p)$, and k is the coefficient of proportionality.

Information and Vision: Summary

Based on my analysis of *Information and Vision*, various components of Glezer's 1961 model of the visual process can be summarized as illustrated in Figure 1 and in the list that follows.

1. **Simultaneous and subsequent contrast**. Perceived differences in luminosity of foreground objects as a result of background luminance and luminance changes can result in distortions of apparent contrast that are not reflected in objective photometric measurements. These contrast-modifying factors can significantly alter perception of images.

2. **Retinal receptive fields in the periphery**. Rod cells over an area of the retina that ranges from 15 arc-min to 1 arc-deg (area of summation under scotopic conditions) collect light for a single ganglion cell (accumulating cell) to form a single receptive field that transmits an image element through the optic channel. This rod field size changes minimally as light levels increase. Thus the smallest detail generally discernible to peripheral vision falls within a 15-arc-min to 1-arc-deg size range.

3. **Retinal receptive fields in the fovea**. Cone cells concentrated in the fovea similarly collect light for a single ganglion cell. The cone receptive field varies in size from about 5 arc-deg under scotopic conditions (200 fields, 90 arc-deg total foveal subtense), to 10 arc-min at mesopic light levels, to 1 arc min at 30 cd/m^2, to the size of a single cone (0.4 arc-min) at about 50 cd/m^2. Thus the smallest discernible detail depends on the light level.

4. **Critical duration**. The critical minimum duration of a light stimulus for a receptive field to respond to it is about 0.1 second at low light levels. The critical duration decreases somewhat as illumination increases. Stabilized images exposed on one spot on the retina for less than the critical duration have better resolution than non-stabilized images. When an image is exposed longer than the critical duration, natural eye movements increase resolution and a stabilized image has poorer resolution.

5. **Temporal accumulation**. A receptive field accumulates photons over time, summing them over the area of accumulation to result in a perceived light level that depends on stimulus size, luminance, and presentation time. Under strong photopic conditions and with good apparent contrast levels, approximately 0.08 second is necessary to perceive that an image is present. The object's general outline can be recognized in about 0.25 second. Identification is possible if the object is presented for about 0.6 second.

6. **Time of inertia**. Perception of the image is maintained in the optic system for 0.012 to 0.2 second after the stimulus disappears for the fovea, and from 0.1 to 0.32 second for peripheral vision. Time of inertia increases as stimulus intensity increases.

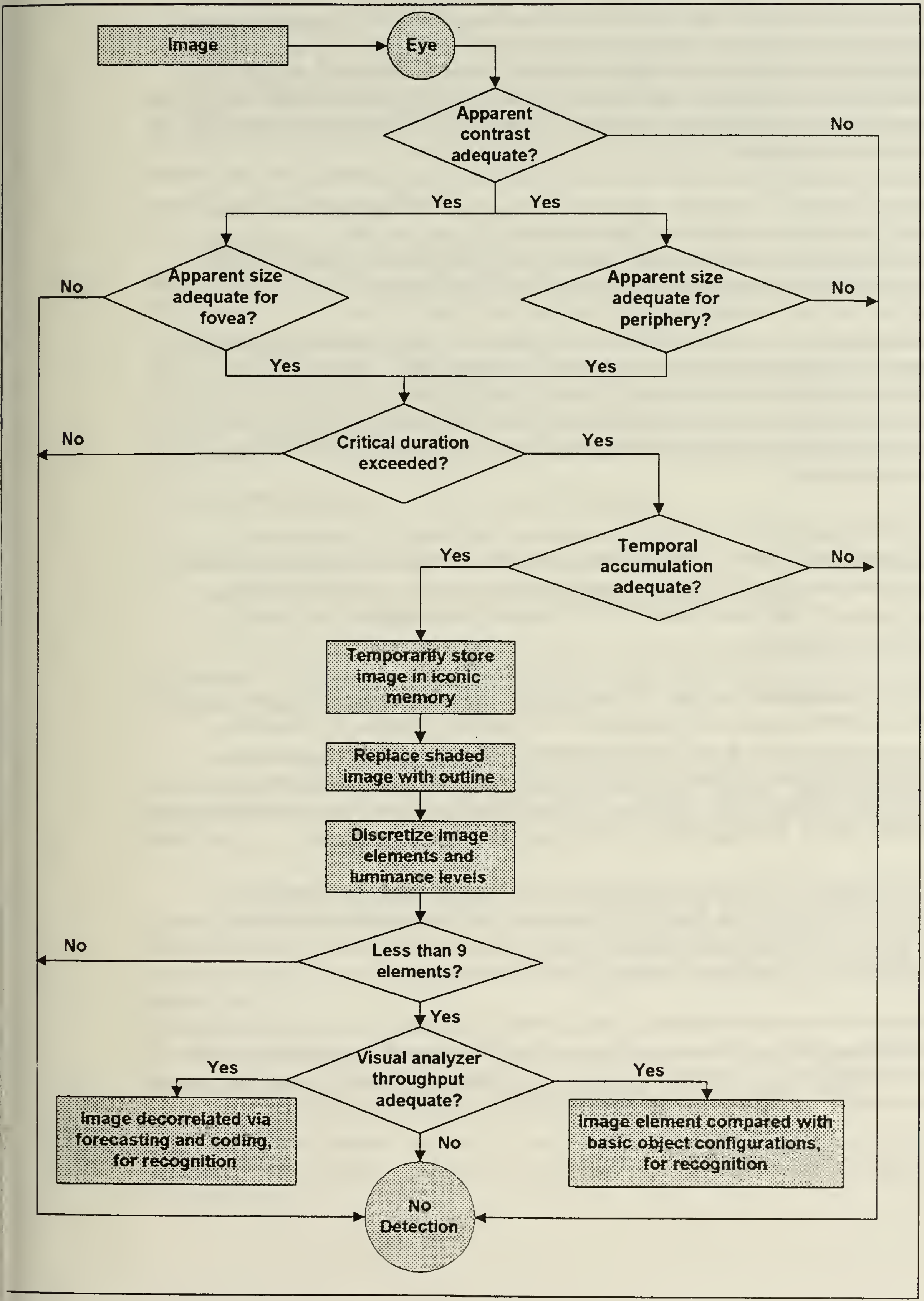
Image
Eye
Apparent contrast adequate?
No
Yes
Yes
Apparent size adequate for fovea?
No
Apparent size adequate for periphery?
No
Yes
Yes
Critical duration exceeded?
No
Yes
Temporal accumulation adequate?
Yes
No
Temporarily store image in iconic memory
Replace shaded image with outline
Discretize image elements and luminance levels
Less than 9 elements?
No
Yes
Visual analyzer throughput adequate?
Yes
Yes
No
Image decorrelated via forecasting and coding, for recognition
Image element compared with basic object configurations, for recognition
No Detection

Figure 1. Components of Glezer's 1961 Model of the Visual Process.

7. **Outlines**. A shaded image can be replaced by an outline of the same object, with areas of transition from black to white (or between shades of gray) marked by borders. The resultant outline will carry essentially the same amount of information as the shaded image but can be transmitted much more efficiently. Emphasized borders enhance apparent contrast of an object. Borders are especially important for enhanced resolution when exposure time is longer than the critical duration and the image is not stabilized to counteract normal eye movements. Borders that change direction sharply or that outline fine details are especially important for drawing visual attention and for discrimination.

8. **Discrete image elements**. Images are segmented and transmitted through the visual system (retina to optic centers of the brain) as a finite collection of discrete *elements*. A continuous two-dimensional image can be considered a distribution of various magnitudes of luminance on a surface, discretized by the visual system into a finite number of elements.

9. **Discrete luminance levels**. Light intensities are segmented and transmitted through the visual system (retina to optic centers of the brain) as a finite number of discrete *luminance levels*. The level of luminosity of the image is perceived at each discrete point that defines the object's length (or height). For low-light-level photopic vision, the number of luminous gradations is about 9. Discernible gradations of brightness can be considered *symbols*, and the collection of symbols is referred to as an *alphabet*.

10. **Number of objects perceived**. The retina can receive on the order of 45 million bits of information at one time. However, after coding and transmission through the optic system, the observer perceives only about 7 ± 2 objects simultaneously over a 0.1-second period of time. Information is stored in short-term memory for 0.27 second, then disappears (and usually is replaced with other information).

11. **Visual analyzer throughput capacity**. At low illumination levels, throughput capacity grows linearly with logarithmic increase of illumination. An illumination increase of 2 results in throughput capacity growth of about 10 bits per second. At higher levels of illumination, throughput capacity becomes constant, but depends on the task. Individual letters can be recognized at rates of about 70 bits per second. Image recognition can occur at throughput rates of 50 to 70 bits per second. Reading speed is measured at 30 to 40 bits per second.

12. **Image decorrelation**. Certain receptive fields appear to separate signal from noise by accumulating signal strength, then these fields decorrelate the image (decrease interactions between elements) by forecasting (predicting whether one element is followed by an identical or a different element) prior to codifying the image element by element.

13. **Configuration-sensitive receptive fields**. Other receptive fields appear to be sensitive to specific simple configurations (shapes) that make up images. These fields codify a specific shape as an entire simple configuration, not element by element. The alphabet of shapes for which this process occurs is not inherent. It is acquired on an individual basis based on life experiences.

Information Processing in the Visual System

Introduction

Information Processing in the Visual System is the title of the Proceedings of the *IV Symposium on Sensory System Physiology*, held in Leningrad in 1976. Contributors from throughout the USSR provided papers, including V.D. Glezer and K.N. Dudkin. In addition, papers were presented by scientists from Great Britain and the United States. As with the other documents that have been summarized for this report, only those papers and points that may relate directly to human target acquisition have been noted here.

There is some question whether the concepts and findings of the non-Eastern European physiologists and psychologists should be included in the Soviet concepts. For the present, some are being included. The work of these individuals was respected enough so they were invited to the conference. Thus their ideas probably have influenced Soviet thinking on visual perception and the target acquisition process. Also, the date of the conference (between Glezer's 1961 seminal work on vision and the major Glezer and Dudkin works in 1985) increases the probability that these non-USSR papers may have influenced the Soviet school as its models continued to develop — as did documented research results obtained throughout the world that are cited in Glezer's and Dudkin's books.

Accommodation, Saccades, and Fixations

On the Functional Rearrangement of Accommodation with a View to Vision Optimization

V.V. Volkov, Ophthalmology Dept., S.M. Kirov Military Medical Academy

The visual accommodation apparatus helps provide regulation and processing of visual information. The appearance of indistinctly visible objects in space stimulates accommodation, especially if accompanied by chromatic aberration.

Accommodation levels are primarily determined by the stimulus and its characteristics. A low-intensity, poorly-structured stimulus has little accommodation “interest” and is rapidly extinguishable. The volume of accommodation enlarges at presentation of more complex objects and is maximal with letter tests (e.g., with Snellen visual acuity charts).

Accommodation amplitude (diopters) and strength depend on general fatigue as well as other factors. For pilots experiencing high workloads, when accommodation is strong, amplitude tends to remain the same or even increase when workload is reduced. When accommodation is intermediate, amplitude reduces when load drops. When accommodation is weak, amplitude decreases markedly with reduced load.

The Optical Link in the Information Processing System and Visual Recognition

Y.S. Rosenbloom and others, Helmholtz Research Institute of Eye Diseases, Moscow

The relation of pattern recognition to visual accommodation and refraction was studied. Within the range of accommodation, recognition depends only on the angular size of the pattern, not on the viewing distance.

Outside the range of accommodation, the threshold distance of pattern recognition is a linear function of the log of the angular size of the pattern, when patterns range from 2 arc-deg to 10 arc-min. Decreasing size from 10 to 2 arc-min, the threshold distance of recognition is constant. These results indicate that accommodation is involved in pattern recognition, and that 10 arc-min is the critical size that triggers the accommodation reflex that will focus the image on the retina.

Varying the Magnitude of Visual Feedback as a Method of Visual Perception Research

N. Ju and others, Institute of Psychology, Academy of Sciences, Moscow.

Fixation accuracy varies within a range of 3 to 4 arc-min to 2 to 5 arc-deg, depending on the amount of voluntary control, type of task (approximate or precise fixation required), and configuration of objects in the visual field. The eye also may require two to four jumps to aim itself.

The Role of Micromovements of the Eye in Contrast Sensitivity

U.T. Keesey, University of Wisconsin, USA

Small, continual movements of the eye include saccades (amplitude 5 arc-min, duration 25 milliseconds), drift (amplitude 5 arc-min, duration 100 milliseconds), and tremor (amplitude 20 arc-sec, 100 times per second). The roles of these movements include the following.

- Temporal luminance changes produced by drift are responsible for continuous object visibility.
- Acuity and general contrast discrimination are not significantly influenced by image motion. But transient activity at the onset and offset of objects enhances sensitivity to low spatial frequency content.
- Sensitivity in the fovea appears to be controlled by movement of the object image over the retina.

Visual Search and Visual Attention

George Sperrling and M.J. Melchner, Bell Laboratories, Murray Hill, NJ, USA

In normal visual search, the observer searches an array of objects for a critical object by moving his eyes over the array. The pattern of eye movements complicates the analysis of the attention factor. Eye movements can be eliminated by having the observer fixate on the center of the display while presenting a new stimulus (set of characters to be scanned for a specified target character) for fixed periods of time. When the display is shown for 240 milliseconds, this results in approximately the movement sequence that the eyes produce for themselves in natural visual search.

Under these conditions, several results and conclusions should be noted.

- When an observer detects a target, he simultaneously detects both its location and its identity (class).
- The maximum number of characters that can be scanned in an array is about 15 to 25. Increasing the number beyond this does not improve performance.
- Observers approach their asymptotic performance when arrays are presented every 120 milliseconds. Increasing time of presentation to 240 milliseconds improves performance only slightly. Increases beyond 240 milliseconds are of no benefit. As a corollary, when an observer

searches large arrays naturally by means of eye movements (corresponding to a new input every 240 milliseconds), his processing capacity is unused for about half of the time (between 120 and 240 milliseconds).

- The most efficient search occurs when new arrays are presented every 40 to 50 milliseconds (corresponding to 20 to 25 fresh arrays per second). Scan rates in excess of 100 characters per second are achieved by most observers.

The point also is made that observers cannot search simultaneously for a large and a small target as well as they can for equal-sized targets. When instructed to do so, they switch attention from trial to trial between the two sizes.

Information Transmission

Information Capacity of the Eye

A.V. Luisov and N.S. Fedorova (no affiliation listed)

A 10-million-fold drop in luminance causes a mere 60-fold drop in visual acuity, a decrease in the quantity of information transmitted by a factor of 70, and a 280-fold drop in the capacity of the eye. Specific information flow increases rapidly as luminance falls. That is, the less light there is, the more complete its use as an information carrier. The information capacity of the retinal periphery is millions of bits per second, but that of the overall vision system is only approximately 70 bits per second.

Perception of Form and Texture

The Perception of Visual Form*:* a Two-Dimensional Analysis

A.P. Ginsburg, Physiological Laboratory, Cambridge, England

The visual system discards a great deal of information at early stages of processing. Data reduction is carried out via filters. The patterns that remain after certain stages of data reduction account for a variety of perceptual phenomena ranging in complexity from alphabet letters to faces. Frequency filter analysis provides a parsimonious and quantifiable metric for visual information processing in humans.

Each receptive field of a visual neuron is viewed as a spatial filter capable of processing three independent parameters of an object. These are spatial frequency (size), phase, and orientation.

As one walks toward a far-away object, it changes from a speck to an object that can be classified and then identified. During this process the pattern information that has been invariant is the low spatial frequency of the object — the coarse details that provide basic form information. The lower three spatial harmonics can provide sufficient information for classification of patterns.

Studies of Spatial Frequency Filters in the Visual Cortex*:* Characteristics and Structural Organization

V.D. Glezer and others, Pavlov Institute of Physiology, Academy of Science of the USSR, Leningrad

Complex receptive fields respond to a striped stimulus in an optimal way. That is, they are tuned to gratings of various spatial frequencies. The receptive fields appear to be narrow-band spatial frequency filters. The mean bandwidth at half-amplitude is 2 octaves. The maximum of spatial frequencies is in the region of 0.3 to 5.0 cycles per arc-deg.

Study of Spatial Frequency Filtration of Moving Stimuli in the Human Visual System

K.N. Dudkin and V.E. Gauzel'man, Pavlov Institute of Physiology, Academy of Science of the USSR, Leningrad

Narrow-band channels (filters) of spatial frequencies exist in human visual systems. Observers were shown either moving right-angled lattices of various spatial frequencies or light bands of various widths.

For the lattice stimulus, observers were told to determine the detection threshold for the periodicity of the stimulating lattice and also the detection threshold for the borders of the lattice. Stimulus speed ranged from 0.35 to 75 arc-deg per second. It was found that at low speeds the maximum contrast sensitivity is found in the high frequency range and at high speed the maximum sensitivity is shifted to a lower frequency.

The light bands were presented either focused or defocused. It was found that contrast sensitivity depends on the rate of movement of the light band, with contrast sensitivity decreasing as movement rate increases, especially for narrow bands. Defocusing the band decreases contrast sensitivity for narrow bands moving at low velocities; the brighter the band, the greater the decrease in sensitivity. For high rates of movement, changes in contrast sensitivity are insignificant.

Mechanisms of Visual Recognition in Dogs at Different Stages of their Ontogenesis

T.A. Mering and N.V. Prasdnikova, Brain Institute, Pavlov Institute of Physiology, USSR Academy of Sciences, Moscow

Psychophysical investigations of visual stimulus recognition in humans have revealed three types of identification (also demonstrated in dogs):

- Comparison against an innate standard (e.g., estimation of the size of an image or its position and orientation, recognition of grating orientation, or recognition of some simple shapes).

- Classification within a definite alphabet of stimuli (picking out one shape from among others).

- Comparison against a learned standard after training or experience (identification of specifically-trained shapes).

It appears that the identification of most shapes is learned. However, the ability to identify gratings appears at a certain age, without previous training.

Threshold Time for Exposure and Recognition Time of Topical Images

V.M. Krol' and L.I. Tanengol'ts, Institute of Management Problems, USSR Academy of Sciences, Moscow

For an observer to recognize which one of an alphabet of eight similar images has been displayed, the image must be shown (exposed) to the observer for a mininum of 36 ± 0.8 milliseconds. The observer then can continue processing and can classify the image as to its type (results based on 400 measurements).

The entire recognition process takes considerably longer. When simple motor reaction time is subtracted from measured response times, the upper boundary was found to be 414 ± 17.0 milliseconds (400 measurements). The lower boundary was 226 ± 15.6 milliseconds (400 measurements). Thus minimum image exposure time must be on the order of 20% of the duration of the entire process of visual recognition, and 80% of the total time is spent in follow-on processing for which presence of the image is not required.

Recognition of Rotated Images by Man

A.A. Nevskya, , Pavlov Institute of Physiology, Academy of Science of the USSR, Leningrad

When an unusual but well-learned image is rotated in small steps (5 to 10 degrees per step), accuracy of recognition is not affected up to 10 to 15 degrees from the learned orientation. With larger changes in orientation, there is an approximately linear dependence between the angle of rotation and probability of correct recognition. If orientation is changed by as much as 30 to 60 degrees, results vary for various figures and various observers. With presentation times of 100 to 150 milliseconds, many figures could not be recognized at all. Given long presentation times, all figures could be identified correctly.

Reaction time was measured for responding that a second shape is the same as or different from a first, when sometimes the second figure was the same as the first but rotated between 7.5 and 120 degrees. Exposure time was 100 to 300 milliseconds. If the figures were rotated no more than 15 degrees, correct same-different responses generally were given in 300 to 450 milliseconds. For large differences in orientation, response times grew significantly, so that rotations of 120 degrees required between about 500 and 650 milliseconds for figure recognition.

Perception of Size

On Coding of Retinal Size by Visual Neurons

Veijo Virsu, General Psychology Dept., University of Helsinki, Finland

Four conditions decrease perceived size and increase threshold size for evoking maximal ganglion responses.

- Dark adaptation.
- Change from cone to rod vision
- Flicker (temporal modulation).
- Short stimulus presentation time.

The One-Dimensional Nature of Spatial Summation

J.P. Thomas, Psychology Dept., University of California Los Angeles, USA

Spatial summation is one-dimensional when photopic foveal viewing is used. In general, visibility of a solid rectangle of light, shown against approximately a 1,000-cd/m^2 background, increases when the shorter dimension of the stimulus is held constant and the longer dimension increases.

When the width is 5 arc-min in size, there is a monotonic increase in visibility as length increases from 5 to 50 arc-min. The same monotonic increase occurs, but with shallower slope, when the width is 10 arc-min. When the width is 20 arc-min, visibility increases as length increases up to 40 arc-min, then decreases slightly as length increases further to 50 arc-min. When the width is wider than 20 arc-min, a similar pattern is observed.

When the length is held constant and the width is increased, only between 0 and 5 arc-min does visibility increase. Beyond 5 arc-min there is little systematic relationship between visibility and width increase. The lack of a generally increasing relationship means that visibility cannot be a function either of total perimeter or area *per se*.

Perception of Motion

Mechanisms of Motion Perception in the Human Visual System

J. Allik and others, Dept. of Psychology, Tartu State University, and Faculty of Psychology, Moscow State University

Perception of motion depends on several processes or factors.

- The visual low-pass filter detects transient changes in luminance.

- Movement detectors perform correlational analyses of one sequence of luminance changes. That is, each of the two kinds of movement detectors respond to movement in only one direction. Leftward and rightward movements are analyzed by different elements or by different neuronal nets.

- Observation that one direction of movement is occurring inhibits determination that the direction has changed. Approximately 10.7 milliseconds is required to respond to leftward movement if the observer is adapted to rightward movement.

Ocular Following Reflex from Periphery of a Cat's Retina

A. Michalski and M. Kossut, Neurophysiology Dept., Nencki Institute of Experimental Biology, Warsaw, Poland

The receptive field used by the cat for vertical visual pursuit of moving objects has the shape of a horizontal strip approximately 30 arc-deg wide and extending at least 40 arc-deg in the nasal and temporal direction.

Extrapolation of Lines

Visual Extrapolation of Straight Lines

N. Yakimoff and others, Institute of Physiology, Bulgarian Academy of Sciences, Sofia, Bulgaria

The ability of humans to determine the point of intersection between a straight line and the visually extrapolated prolongation of a straight segment has been studied. The angle between the line and the segment was 30, 60, or 90 degrees. The set of imaginary prolongations was found to be displaced for the two acute angles of 30 and 60 degrees, as if these angles had been overestimated.

Visual Identification and Its Neurophysiological Mechanisms

Introduction

Visual Identification and Its Neurophysiological Mechanisms was a joint effort by V.D. Glezer and K.N. Dudkin in 1975, when both were at the Pavlov Institute. The copy available for review is an unedited machine translation, unfortunately, and there is some uncertainty that the authors' concepts and explanations have been correctly interpreted. As a result, considerably more weight is being given to Glezer's 1961 and 1985 books and to Dudkin's 1985 book, during this search for a Soviet model of target acquisition. A reasonably detailed review of the text indicates that little is included here that is not presented in Glezer's and Dudkin's other books.

Types of Visual Identifications of Images

Three types of image or pattern recognition processes have been identified:

- Recognition of simple three-dimensional indicators (criteria) of images with the aid of *inherent* specialized detectors, based on "standard" images. Image properties include line or figure orientation and image location, regardless of image complexity. The time required to recognize, using this process, does not depend on how many symbols or items are in the alphabet (the set of possible symbols).

- Recognition based on selection among items in the alphabet of possible items. Object shapes, geometric figures, and letters are recognized in this way. In the initial stages of training, time to recognize a symbol depends on how many symbols are in the total alphabet.

- Recognition based on a well-trained standard image. Time of recognition does not depend on the size of the alphabet.

Identification Invariance and Visual System Channels

Humans are capable of identifying objects in spite of changes in brightness, contrast, color, retinal location, rotation, and compression or elongation. At the same time, humans can differentiate between variants of images when this is the required task. The authors provide mechanisms that may explain this phenomenon. However, these mechanisms do not seem pertinent to development of a target acquisition model.

Neurophysiological Mechanisms of Visual Identification

The last half of the book is devoted to development of a model that can simulate the vision process *per se*. This does not appear to be directly applicable to the desired model of the target acquisition process.

This page intentionally left blank.

Visual Perception and Memory

Introduction

Visual Perception and Memory was published in 1985. Its author, K.N. Dudkin, has worked in close collaboration with V.D. Glezer at the Pavlov Institute. However, unlike Glezer, Dudkin's visual perception emphasis is more with cognitive than with physiological mechanisms and is based on concepts first developed by I.M. Sechenov. Dudkin holds that in the cortex there is no single universal method for describing perceived objects. Instead, long-term memory and the context in which events occur control streams of impulses through the neuron structures and result in perception (or the lack thereof).

Dudkin attempts to combine the results of psychophysical (humans), behavioral (monkeys), and neurophysiological (cats) experiments within a general framework of models and hypotheses. The mechanisms of visual memory are considered the basis for visual perception. That is, visual perception is most likely determined by a complex hierarchical memory system, based on the integrative activity of the brain, that links the visual system with other mechanisms. Memory is examined not only as a repository of information but also as a system of sensory and control processes organized according to specific rules.

The book includes physiological mechanisms of iconic, short-term, and long-term memory, and the role of these mechanisms in visual perception. Dudkin's approach to a physiological model is based on the concept of parallel elementary units of perception: spatial-frequency filters and object property detectors.

System of Visual Memory

Perception is impossible without memory. As noted by I.M. Sechenov in 1952, "That which is seen and heard by us always includes elements already seen and heard before. In view of this, during any new viewing or hearing, to the products of the latter [viewing and hearing] are attached similar elements which can be reproduced from the memory bank, although not individually but rather in combinations in which they are recorded in the storehouse of memory." Dudkin considers visual perception a specific stage in the process of organizing behavior, based on the diverse mechanisms of memory.

Pattern Recognition

The concept of *image* (also translated as *pattern* or *figure*) can be treated as a generalized description of a group or class of similar objects. Image or pattern recognition can be considered either (1) classification of the perceived the objects (identification of the class to which they belong) or (2) a detailed description of the objects plus the relationships of all their parts.

Recognition can be considered a process of *classification*, where the multidimensional space of indicators is shaped and the hypersurfaces that demarcate kinds of objects are separated into classes. Classification occurs in three stages: (1) sensory coding, (2) indicator (criteria) extraction, and (3) classification based on the isolated indicators. Neurophysiological structures in permanent memory first distinguish the indicators and transform image inputs into a point or a vector that represents the indicators in multidimensional space. Using decision rules, the space of indicators is divided into regions, each region for a different class of objects. Image variants belong to different classes of objects, each with its own set of points or vectors. Recognition of an object occurs when the vector of indicators for an object matches the set for a specific class.

Alternatively, the process of recognition can be considered *structural* (linguistic) in nature, characterized by description of the image with the aid of a dictionary of indicators (signs, criteria) or primary elements. This description includes the relationships among these elements, their hierarchical structure, and their rules of association. A scene represents a statement in the grammar of a given language, with the language determined by the interrelations of the parts of the scene. Analysis of a scene is similar to grammatical analysis of a sentence. The structural approach can be used to explain visual analysis of complex scenes in which interrelationships among fragments are determined by context.

The Problem of Invariance

An object can undergo various spatial transformations, yet still can be recognized. Similarly, humans can discriminate the difference between a transformed image and a variation on that image which now makes it different. This constancy of perception is referred to as *perceptual invariance*. Sechenov suggests that invariance in recognition is provided by general properties of objects, retained during transformations. The process of learning via investigation (a large role in which is played by eye movements) results in formation of neural networks that retain traces of effects and make it possible to isolate an object's properties. New neural associations also can be formed, uniting those that existed earlier and resulting in invariance in discrimination.

Types of Memory

Three types of memory generally are recognized by psychologists, classified on the basis of duration of information storage. These are referred to as iconic, short-term, and long-term memory.

Iconic memory has a duration of a few hundred milliseconds. If not utilized during this time, information is lost forever. The capacity of iconic memory is of the order of 9 letters maximum if reported immediately, dropping to 5 letters after 1 second, and diminishing rapidly thereafter. Information held here is precategorical, that is, not yet recognized and put into a category. It is stored in the visual form as a sensory trace. During the process of recognition the sensory trace is lost and the information is passed along to the following links in the memory system.

Information stored in *short-term memory* is retained for 15 to 30 seconds, with retention time determined by the type of input information. The contents of this type of memory also disappear irreversibly if not utilized. Information held here is postcategorical. Identification is considered a rapid process, realized by automatic mechanisms that distinguish the specific properties of the stimulus. The description of the object is retained as a set of characteristics and properties of the image.

When information is needed over a prolonged period, it is memorized into *long-term memory* or permanent memory. Here information can be stored without loss for an indeterminate period of time. Information is stored in long-term memory using one of three mechanisms. *Address* storage is characterized by inclusion of the coordinates of the storage location along with the information. *Content* storage occurs when the location in memory is not stored, but instead some “key” is stored with the information when it is written; the key is used to evoke the appropriate output when reading of the information is needed. *Associative* storage includes both key information and also context information that is used to write it to the proper depository and to read it from there.

Short-Term Memory

Dudkin suggests that sensory data is processed for temporary storage in short-term memory via a sequence of operations. First the data are coded in iconic memory, where the image (or its description) can be stored for 0.25 to 1.0 seconds. The information then is read into the short-term memory's *recognition buffer* at about 100 symbols per second. There the data are converted from visual form to a program of *motor instructions* and stored until they are moved to the *repetition block*. The process of repetition provides prolonged retention in short-term memory and makes possible the conversion of the trace to long-term memory.

The results of recognition are converted into codes and stored in verbal short-term memory. These results can be transmitted via the *repetition block - verbal depository* loop prior to implementation of the motor program. Visual coding is considered *rapid* coding of information into short-term memory, and occurs using scanning to move sensory data to the recognition buffer. Semantic coding and the attachment of names to objects results in *slow* coding; this process converts the results of recognition into codes that are stored in verbal short-term memory.

There is a bilateral exchange of information between short-term and long-term memory. *Control processes* are necessary to call up information from a specific address in long-term memory for use in short-term memory during coding.

Visual Long-Term Memory

Sechenov provides a model of long-term memory that includes everything that the individual knows about the world, stored in the nervous system. He notes that impressions of objects and their characteristics, qualities, states, and interdependencies are entered in storage as (1) what preceded this impression, (2) what accompanied it, (3) what followed it, and (4) with what it is similar, in whole or in part.

Based on information in long-term memory, objects are detected and identified, figures are discerned from background, and images are recognized. These operations are carried out in the working buffer of short-term memory via the comparison of sensory information with data recalled from long-term memory. Long-term memory apparently contains standards for and prototypes of objects. These include indicators that are characteristics of images, images themselves, and cognitive maps or charts. The type of code used depends on the method of recognition, which apparently is determined by the type of information perceived.

Standards

It is possible that recognition is achieved by comparing sensory information with *standards* held in long-term memory. These standards can serve as mechanisms for automatically distinguishing properties such as the orientation of boundaries and outlines, line lengths, the position of an object in the field of vision, and its texture. Inherent mechanisms apparently exist that automatically distinguish lines of different orientations, location in the field of vision, the distance of objects, object luminance values, and motion.

Neuron structures which form receptive fields have weighting functions that determine their spatial frequency properties. These weighting functions are quantitative measures of excitatory and inhibitory neuron connections formed in long-term memory. Various types of automatic mechanisms are used for spatial frequency filtering. Some of these apparently are matched filter standards used to search for specific properties of an object, using cross correlation of the standard with the observed image. Such matched filtering probably is used during analysis of textures.

With prolonged training, humans apparently can develop standards they can use for recognition, even for complicated images. However, Dudkin notes that, if standards are the primary mechanism for recognition, an extremely large volume is required in long-term memory since it is necessary to memorize standards for all objects.

Indicators, Images, Prototypes

Objects subjected to noise and distortions are more difficult to recognize. During comparison of sensory information with information in long-term memory, approximate similarity is checked based on distinguishing properties and characteristics common to many objects.

Recognition implies the ability to discern from the total sum of indicators those that are most characteristic, and to memorize this group. The problem of recognition is to discern a set of indicators (criteria) that describe an object and to compare the object with the generalized description that is stored in memory. A generalized description of a group of similar objects can be called an *image*. That is, an image often is understood to be a certain generalized, idealized schema or prototype that includes a set of rules for composing its description.

Indicators, primary elements, and prototypes are selected based on the *context* in which the object is embedded. This simplifies the search in long-term memory, since it is possible to restrict the number of images that correspond to a given object.

Using the *classification* model of pattern recognition, the image is stored in the form of a multidimensional vector of indicators. The process of recognition is reduced to classification of objects. That is, recognition consists of picking out the appropriate regions of the space of indicators, with the aid of specific separating functions (functions based on indicators that characterize the differences between images of different classes).

In the classification model, long-term memory possibly contains various systems or sets of indicators. Which is selected is determined by the type of perceived sensory information and by the method of recognition. In Dudkin's view, this model has advantages over the *standards* model in that less storage capacity is required and invariance to transformations is explained.

The *structural* model of pattern recognition uses an approach based on symbolic structural description of objects — idealized schema that include sets of primary elements and their grammars and that assign rules of grammatical selection during analysis. The specific sets of primary elements depend on the type of perceived sensory data. Variants are formed by transforming an object, that is, by changing its dimensions, rotating it, or distorting it with noise. Variants are described based on the prototype that is stored in memory. Normalizing operators are necessary for processing sensory information, in order to explain invariant discrimination.

Cognitive Maps and the Theory of Frames

Dudkin assumes that items that are stored in long-term memory include *mental models* or *cognitive maps,* that is, generalized idealized diagrams based on the processes of perception, collection, and/or synthesis of sensory information. The diagram is specific to what is perceived and can be modified by experience. It accepts information, and changes as a result of this information. It guides motions and investigative activities that enable access to new information which in turn further alters the diagram. The cognitive map takes an active, organizing role in the collection of sensory information and in comparing it with what is stored. Context is very important in this process.

Minsky's theory of *frames*, based on research in the field of artificial intelligence, is similar to the concept of cognitive maps. The frame is a generalized system of data stored in long-term memory; it contains various kinds of information about frequently encountered scenes and situations. Each frame contains information about the situation in which it is to be used, the results to be expected from the perception, and what to change in case of unconfirmed results. It also may contain data needed for the specific perceived scene or situation, data that can be replaced with information contained in the scene.

Interrelationships in the Visual Memory System

The system of visual memory is a complicated information structure with bi-directional interconnections among its separate components. The memory system structure changes in the process of information accumulation. The structural components are the basic depositories into which the required data are recorded and from which these data are read out. Control processes are used for data processing. These processes are critical for coding of information, information repetition so it can be retained longer, and search for required addresses of depositories.

In Dudkin's model, information passes from the iconic memory buffer to short-term memory, which includes the working buffer used for pattern recognition. From there information can be passed to long-term memory. In turn, long-term memory passes appropriate information to short-term memory for correlation with what is perceived. Control processes coordinate information flow throughout the overall system.

Dudkin regards memory as a complex form of brain activity realized by a hierarchical system that includes a set of neural structures. The result of this activity is the ability to process, store, and reproduce perceived information. Most likely the brain contains a large quantity of specialized depositories into which (after appropriate processing and coding) perceived data are distributed. Depositories are organized so that information can be both recorded in and called up from them.

Role of Brain Hemisphere Asymmetry

Dudkin points out that it is well known that the right cerebral hemisphere is better adapted for spatial processing and the left for linguistic and analytical problems Information presented to the left or right visual field enters into the right or left hemisphere, respectively. Simple images (well-known objects) are processed equally well by either hemisphere. Complicated images are better interpreted by the right hemisphere; if detailed analysis is needed to recognize the human face or photos, the right hemisphere will perform the task better. However, if detailed analysis is not needed, the left hemisphere will recognize these images more easily.

Thus Dudkin proposes that the two hemispheres have different functions in image recognition. The left serves for schematic, generalized (invariant) descriptions, while the right is better for specific, detailed description of objects. On the basis of these results, he suggests that the left hemisphere uses the *classification* model of pattern recognition, characterized by initial invariance. The right hemisphere then uses the *structural* model, characterized by secondary invariance modified through the process of learning.

Models of Visual Detectors and Spatial Frequency Filters

Four main cognitive processes relate to the recognition of images. First, information must be gathered from the spatial-temporal luminous flux which reflects and maps the properties of the external world. Next, the information must be organized in a certain manner. Then the information must be compared with information held in memory. Finally, the obtained information must be recorded into long-term memory depositories during learning.

Possible Mechanisms for Describing Visual Objects

Indicators (properties of an object that are available to the senses) represent products of separate physiological reactions of perception. The number of the former is strictly determined by the number of the latter.

All neurophysical data concerning extracting indicators can be subdivided into two groups, leading to two main models: detectors and spatial frequency filters. Each of these models includes properties and characteristics of objects or indicators of them. Apparently, one single, universal method is not used by visual neuronal structures for processing and enscribing all sensory information. Therefore models that hypothesize spatial frequency filters and those that propose a process involving form-specific detectors should not be considered mutually exclusive, but rather appear to complement each other.

Property Detector Model and Ecological Approach

The behavior of living organisms is determined by their connections with the surrounding world. Perception and recognition of those properties of the world that are important for organizing behavior are the responsibilities of the visual system of all animals, including man. This is the *property detector model* or *ecological approach* for perception and behavior. *Detectors* are assumed to be neuron mechanisms that store standards for the ecological properties of visual objects.

Detectors of various image properties are found in many species of animals. Seven types of visual mechanisms or detectors have been proposed for human visual analysis:

- Perception and measurement of brightness and color.
- Perception of outline and gradient.
- Perception and measurement of curvature.
- Perception of angles, intersections, and breaks in lines and boundaries.
- Perception of spots.
- Perception of texture (surface structure) and measurement of the texture gradient.
- Detection and measurement of speed and acceleration.

Spatial Frequency Filter Model and Fourier Analysis of Images

The *spatial frequency filter model* of vision proposes that the distribution of illumination (intensity) can be represented as a function based on superimpositions of harmonic components whose amplitudes are the Fourier coefficients of the function's spectrum. The spectrum is discrete for periodic functions (arrays of harmonic components whose frequencies are multiples of the fundamental frequency). The spectrum is continuous for aperiodic functions.

Research indicates that contrast sensitivity is determined in a rather broad band by the amplitude of the fundamental harmonic. A low-contrast grating first is detected as sinusoidal. Greater contrast then permits recognition of its wave shape (e.g., as a square wave grating). Complex gratings are perceived as other than sinusoidal only if the detection threshold has been reached for upper harmonic components.

Dudkin proposes that the visual system includes a set of independent channels, each with bandwidth between 0.5 and 1 octave and each tuned to a narrow spatial frequency range. The output from each channel is a threshold device for detection of a specific signal. Amplification in each channel does not depend on other Fourier components that describe the image's spectrum, and the visual system can be considered linear.

An optical model of the Fourier transformation of visual information has been proposed. According to this model, form perception is the result of spatial frequency analysis during which low-frequency components of the spectrum describe the object itself and high-frequency components describe the elements of the object. Very high spatial frequency components are necessary for pattern recognition. Using photographic images of the human face, basic information about the object can be obtained with the aid of four low-frequency filters with spatial frequency tuning of 2, 4, 8, and 16 periods (cycles) per image width.

That is, simple and complex visual cortex receptive fields perform piecewise generalized Fourier transforms of images. This is done by breaking the images down using a specific system of basic orthogonal functions. The result of the generalized Fourier transform is a set of *coefficients of expansion* in terms of the baseline functions that form the space of indicators (criteria). This set then represents a point in space and can be used to classify images.

Filters possibly assist with several functions during the preattentional first stage of processing: (1) extract the figure from the background, (2) segment the images, (3) distinguish boundaries, and (4) separate out textural regions with different spatial frequency spectra.

Perception of Periodic and Aperiodic Visual Stimuli

Results obtained while studying contrast sensitivity for sinusoidal gratings, light bars, and edges (contrast boundaries) have led to the conclusion that several different types of detectors exist: gratings, bars, and edges. Grating detectors are narrow-band filters. Bar and edge detectors are wide-band low-frequency filters. The ecological approach to perception assumes that these object properties, along with texture, play the most important role in behavior organization in man and other animals.

Sufficient grounds have not been developed to contradict either the visual model based on detectors or that based on spatial frequency filters. Most likely they complement each other. As noted by A.A. Kharkevich, the selection of one or the other method of describing the system depends not so much on the system's arrangement as on the model's purpose and use. System properties do not change with the method used to describe them. The unconditional application of the spectral method is not wise.

Receptive Field Linearity and Nonlinearity

Each receptive field is considered to be a specific operator that acts on the original image to convert it to signals. Conversions can be linear or nonlinear. The receptive field often is modeled as a linear spatial frequency filter, and Dudkin's model requires acceptance of the hypothesis that the receptive field meets the requirements for a linear filter. However, experimental results so far are inconclusive.

For *linear conversions*, when separate signals $s_1(x, y)$ and $s_2(x, y)$ are projected on the retina, each must stimulate responses $r_1(x, y)$ and $r_2(x, y)$. Linear combination of the stimuli $a_1 * s_1(x ,y) + a_2 * s_2(x, y)$ must result in the response $r(x, y) = a_1 * r_1(x, y) + a_2 * r_2(x, y)$, that is, linear superimposition of responses based on the input signal. This is a necessary requirement for linear systems. A receptive field can be modeled as a *linear three-dimensional invariant filter* only (1) if it meets the principle of superimposition and (2) if, during excitation due to the signal $s(x, y)$ with response $r(x, y)$, the shifted signal $s(x - x_0, y - y_0)$ yields a shifted but analogous response $r(x - x_0, y - y_0)$.

Sinusoidal gratings can be used as input stimuli for different receptive fields. This makes it possible to determine for all spatial frequencies the transmission characteristics (amplitude and phase) of the linear spatial frequency filters (receptive fields) that assign output amplitude and phase ratios to input amplitude and phase ratios. When the transmission characteristics have been determined for all spatial frequencies, a filter's response to any nonsinusoidal stimulus then can be predicted, since any input image (signal) can be represented by the superimposition of the individual harmonic components of the signal.

When the receptive field is modeled as a linear three-dimensional invariant filter, responses to input functions (that is, the distribution of illumination in the field of vision) can be determined through the integral of convolution. However, a convolution operation in the spatial domain is equivalent to multiplication in the frequency domain. If Fourier transforms of the functions are carried out, then $R(f_x, f_y) = H(f_x, f_y) * S(f_x, f_y)$. The functions $R(f_x, f_y)$, $H(f_x, f_y)$, and $S(f_x, f_y)$ are the Fourier transforms of $s(x, y)$, $r(x, y)$, and $h(x, y)$. The last function $h(x, y)$ is a weighting function which depends on the system's response to the input signal. The Fourier transform $H(f_x, f_y)$ of the weighting function is the transmission or transfer characteristic which determines the spatial frequency filtering properties of visual mechanism. With linear systems, no frequencies are included in output that were not part of the input spectrum.

For *nonlinear conversions*, separate signals $s_1(x, y)$ and $s_2(x, y)$ also cause responses $r_1(x, y)$ and $r_2(x, y)$ but the linear combination $a_1 * s_1(x, y) + a_2 * s_2(x, y)$ results in a different total response $r(x, y) = \varphi * [a_1 * r_1(x, y) + a_2 * r_2(x, y)]$ that is not a linear superimposition of the responses of the separate signal components. Thus these are considered to be nonlinear systems.

The inapplicability of the principle of superimposition is a basic property of nonlinear systems. To analyze such systems, responses to complex input signals must be investigated without breaking the signals into simple components. New frequencies can appear in the output, even with simple harmonic signal inputs. The composition of the output signal's spectrum depends on the input signal's form and amplitude. For the visual system, this process is very difficult to model.

Basis of Iconic and Short-Term Memory

The retention of information during visual perception (necessary for pattern recognition and behavior organization) appears to be realized at various neuron levels. Dudkin suggests two mechanisms of vision, tonal and phasic, to help describe the retention of information. These mechanisms (or something similar) probably are critical to the storage of information in iconic memory.

Tonal and Phasic Mechanisms

A *tonal* (sustained) mechanism is sensitive to the gradients of illumination in space and reacts to the form of objects in the field of vision. This is a rapid-response mechanism, reacting to high temporal frequencies of stimulus on-off, and responding with continuous activity to a moving grating. Tonal neurons process high-frequency three-dimensional patterns, retain information longer, and have a greater time constant for development of reactions.

A *phasic* (transient) mechanism is sensitive to changes in illumination over time and, to a much lesser extent, in space. This mechanism detects temporal contrasts, boundaries, and object motion. It demonstrates continuous activity as stimuli switch on and off, and a modulated response to motion of a grating through the receptive field. Phasic neurons respond to higher temporal frequencies and lower spatial frequencies than do tonal neurons, and response latency is shorter.

A new object that appears in the visual periphery is first received by the phasic neurons of that portion of the brain which controls head rotation and eye movements; physical movements that result in the projection of the object onto the fovea thus occur. Phasic mechanisms precede tonal mechanisms by 50 to 100 milliseconds. The phasic neurons at this stage inhibit the tonal neurons, resulting in saccadic suppression during eye motion. Integration and retention of information in the tonal system begins during the following fixation. Thus different properties of the image are collected in iconic memory by the two mechanisms. The phasic system conveys information about object location or rapid location changes, while the tonal system conveys information about object form.

Detection of Periodicities

Research indicates that recognition of moving periodic gratings is a function of the speed of grating motion and the spatial frequency of the grating. Beyond a certain speed range, the ability to perceive the gratings is seriously hindered. For high frequency gratings, this breakpoint occurs at a relatively slow rate of motion. That is, contrast sensitivity (the reciprocal of luminance threshold) for higher spatial frequencies drops with higher rates of motion. For low frequency gratings, higher speeds are better for perception. These experimental results agree with the differences in spatial and temporal characteristics of the tonal and phasic mechanisms.

Detection of Grating Boundaries

As gratings are moved along a surface, contrast sensitivity for the boundaries of these gratings (the aperture into which the gratings are inscribed) also can be measured. Contrast sensitivity is considerably lower for these boundaries than for the gratings themselves, with the greatest difference for a stationary grating. As velocity increases, contrast sensitivity for high spatial frequency gratings falls rapidly while sensitivity for the boundaries stays nearly constant. Boundaries are perceived as blurred during motion, with blurring greater at higher speeds. However, even at high speeds when periodicity of the ratings cannot be recognized, the boundaries are easily perceived, though they appear blurred.

Detection of Single Light Bars

A single vertical light bar, varying in width from 0.116 to 5.7 arc-deg, was moved across a background with luminance of 2.4 cd/m^2 and the observer's contrast sensitivity was determined Sensitivity to contrast levels depends on the rate of motion of the bars. Greater rates of motion and narrower bars resulted in greater reduction in contrast sensitivity. If the resolution of slowly-moving narrow bars is defocused (leading to a Gaussian illumination profile), contrast sensitivity decreases markedly.

Effect of Spatial Frequency Spectrum on Short-Term Memory

The retention of information in short-term memory depends substantially on its spatial frequency content. High-frequency information (contrast boundaries, outlines, fine details, some fine-grained textures) is retained up to 10 times longer than low frequency information (large regions of uniform brightness, smooth illumination gradients, low-frequency textures).

Separation of Figures from Background

Objects in the visible world are perceived by humans as figures arranged against a background. Characteristics of the environment and internal human mechanisms determine which objects are seen as figures and which as background.

The task of the visual mechanism in separating figure from background consists of segmentation of the visual scene. The process includes detection of the boundaries of objects that emerge as figures, the breaking out of fields with uniform distributions from those that are textured, and picking out and describing fragments of scenes for use if needed during pattern recognition processes.

Attention Models: Active Perception

Critical to the process of separating figure from background is the concept of *attention*. One model, referred to as the *model of active perception* or the *controlling model*, postulates the existence of two stages in perception: a preattention stage of parallel processing and an attentional stage of series processing.

The *preattention stage* is carried out by *automatic mechanisms*. Processes during this stage separate out the environment's specific properties and characteristics via a set of neuronal structures, either inherent or formed during the process of learning. These automatic mechanisms include both detectors of properties and spatial frequency filters. The mechanisms serve as elements of a system that stores indicators (criteria), standards, prototypes, subimages, and images (patterns) in long-term memory.

The *attentional stage* utilizes *controlling processes* that manipulate information flow to and from memory. These processes include coding, making decisions, and searching short-term and long-term memory. Sensory information is processed and memorized better in one situation than in another as a function of the setting (adjustment) of the perceiver. Attention appears to be selective, so that information from the environment is received piecewise in a specific sequence.

Automatic processes that separate out environment properties do not always precede controlling processes. The two can occur in parallel, or controlling processes can trigger automatic mechanisms during image synthesis. Regions of images that require detailed investigation can be separated out via the controlling processes, and fragments are located that must be combined with others. That is, the controlling processes segment scenes in order to obtain optimum descriptions and to identify indicators needed for pattern recognition. Automatic mechanisms can be retuned when needed to perceive sharply differing objects.

Controlling processes can be restructured or readjusted through training, resulting in formation of new or revision of already available mechanisms. Thus controlling processes make it possible to orient oneself in an unknown environment and to form new information structures in memory.

Attention Models: Passive Perception

A second model, the *model of passive perception* or the *filtering model*, also has some basis in research results. Here the function of attention is reduced to filtering of some properties and the blocking or considerable weakening of others during the first stage of perception. That is, neurons serve as selective filters that pass necessary signals on through and block the rest, using a system of command neurons.

In this model, complex conversions of information occur in the second stage as it passes through various filters. Analyzing elements, which assign values to the determining indicators (criteria), participate in the processing. The stimulus sequentially passes through all stages of processing, with the perceived pattern or image formed in the output. At the highest level of perception, information about the indicators is compared with that previously accumulated in memory. As a result, a conscious understanding of the synthesized image emerges.

The passive perception model suggests that selective inhibition of spatial frequency and orientation filters is the basic principle of the mechanisms that separate figure from ground and identify textures. This process apparently occurs using automatic mechanisms during the preattentive phase of vision.

Perception and Recognition of Texture

The visual system perceives some “objects” in the field of vision as regions that consist of numerous elements (distinguishable brightness gradients) which are randomly or regularly distributed over the area. If these regions are perceived as uniform, they can be considered textures. Dudkin proposes that texture recognition occurs in the preattentive phase of vision as parallel automatic processing of images is taking place.

As yet there is no general approach to the description of textures. Most frequently they are characterized qualitatively: fine- or course-grained, flat or hilly, etc. However, sometimes spatial frequency descriptions are used for textures.

One- or two-dimensional textures can have statistical connections of elements in two-dimensional space. *First-order statistics* characterize the distribution of combined probabilities that a point (a one-dimensional object of various sizes), randomly located on the texture, falls on the black (or white) color. *Second-order statistics* characterize the distribution of combined probabilities that both ends of a line (a two-dimensional object of various sizes and orientations), will fall on the black (or white) color. Naturally-occurring textures can be described using these statistics.

If figure and background can be characterized using first-order statistics, humans can easily and rapidly (160 milliseconds) distinguish the figure. If second-order statistics must be used, the figure usually cannot be separated from the background even though the same point in the distribution is used as for first-order statistical descriptions.

Classes of textures called *texons* have been identified based on quasi-linearity, closure, angle, connectivity, and grain size. Other texons have been postulated: color; elongated “droplets” of specific orientation, width, and length; and density of micropattern. These are considered the basic local indicative elements that make it possible to distinguish textures. Some pairs of textures, even with identical second-order statistics, can be distinguished even after very brief exposure. Apparently the differences in texons facilitate perception, with an increase in the number of texons increasing the probability of perception. Presence of texons in the figure improves perception more than their presence in the background.

Image Storage During Learning

The concept of a *cognitive structure* (also referred to as a standard, image, prototype, or cognitive map) is widely used in cognitive psychology. This implies a certain type of permanent *memory*, in the form of mechanisms and the processes connected with these mechanisms, formed during learning. An important function of the visual system is participation in the formation of these cognitive structures.

Models of Mental Images

Pattern recognition is possible only because cognitive information structures, formed during the learning process, are stored in long-term memory. Impressions obtained during perception of an image and reproduction of that image from memory appear to be identical. An imaginary image appears to differ little from that formed during perception.

Time of reaction to a specific property of a mental image depends on the image's size and complexity. Small size and great complexity both increase reaction time.

Dudkin proposes two models to explain how mental images are stored. One considers that the images are stored in *visual form* in a visual buffer, including spatial metrics and all properties inherent in the actual visual object. It is proposed that the codes formed during synthesis of visual images are radically different from codes for verbal information. The organization of the visual buffer is inborn and is limited in size. Information is represented in the buffer by selective activation of local regions of coordinate space in the cortex.

The process of image formation includes: (1) generation of mental images, (2) checking conformity of the mental image against what is required, (3) transformations of the image (shifting, turning, scale changes), and (4) spontaneous evocation of information from long-term memory. It is assumed that the cortex-level buffer includes a "screen" on which long-term memory information (fragments of objects or scenes) is called up. A seemingly latent image of the mental image is formed on this screen.

The second model assumes that an image is stored in *abstract symbolic form*, without spatial metrics. It can be reproduced by processing and analysis of stored symbolic information. That is, information is represented in a single abstract symbolic format and is understood by means of grammar analysis procedures. The model is based on lists of "statements" (symbols), each of which is given a specific name. Thus the model depends on language, and cannot account for evidence that animals also can perform recognition tasks.

Conclusions

Dudkin concludes his book with a fairly detailed mathematical model of the transmission of information throughout the visual system (not directly applicable to modeling the target acquisition process), based on differential equations. He then summarizes the implications of the extensive work on visual perception reported in the book.

Memory is considered not only a depository of information, but also the totality of sensory and control processes that affect perception. Recognition is considered the simplest form of thought. The processes of perception are based on sensory data, but cannot be realized without the mechanisms of memory. Perceptual processes include

- Extraction of characteristic features from the environment.
- Memorization of these features during the learning process.
- Classification of images, based on detailed descriptions of objects.
- Comprehension and understanding of the images.

Visual Perception and Memory: Summary

Dudkin's model of the visual process deals both with storage of information in memory and with the perception and recognition process. Points related to *information storage* can be summarized as shown in Figure 2 and in the following list.

1. **Memory structures**. Three types of memory systems are used in conjunction with visual perception: *iconic* (duration < 1 second, capacity 9 simple symbols), *short-term* (duration 15 to 30 seconds, capacity 5 to 9 items), and *long-term* (duration indefinite, capacity unknown). Information passes from the iconic memory buffer to short-term memory, which includes the working buffer used for pattern recognition. From the short-term store information can be passed to long-term memory. In turn, long-term memory passes appropriate information to short-term memory for correlation with what is perceived. Control processes coordinate information flow throughout the overall system.

2. **Short-term storage**. Retention of information in short-term memory depends substantially on its spatial frequency content. High-frequency information (contrast boundaries, outlines, fine details, some fine-grained textures) is retained up to 10 times longer than low frequency information (large regions of uniform brightness, smooth illumination gradients, low-frequency textures).

3. **Long-term storage**. Information about objects, including their characteristics, qualities, states, and interdependencies, is entered into storage along with (1) what preceded this impression, (2) what accompanied it, (3) what followed it, and (4) what it is similar, in whole or in part. These “keys” or “handles” assist in information retrieval.

4. **Classification model**. Each image may be stored in long-term memory in the form of a multidimensional vector of indicators. Pattern recognition consists of picking out the appropriate regions of the space of all stored indicators, with the aid of specific separating functions. Long-term memory possibly contains various systems or sets of indicators. Which is selected is determined by the type of perceived sensory information and by the method of recognition.

5. **Structural model**. Images may be stored in long-term memory in the form of symbolic structural descriptions of objects — idealized schema that include sets of primary elements and their grammars and that assign rules of grammatical selection during analysis. The specific sets of primary elements depend on the type of perceived sensory data.

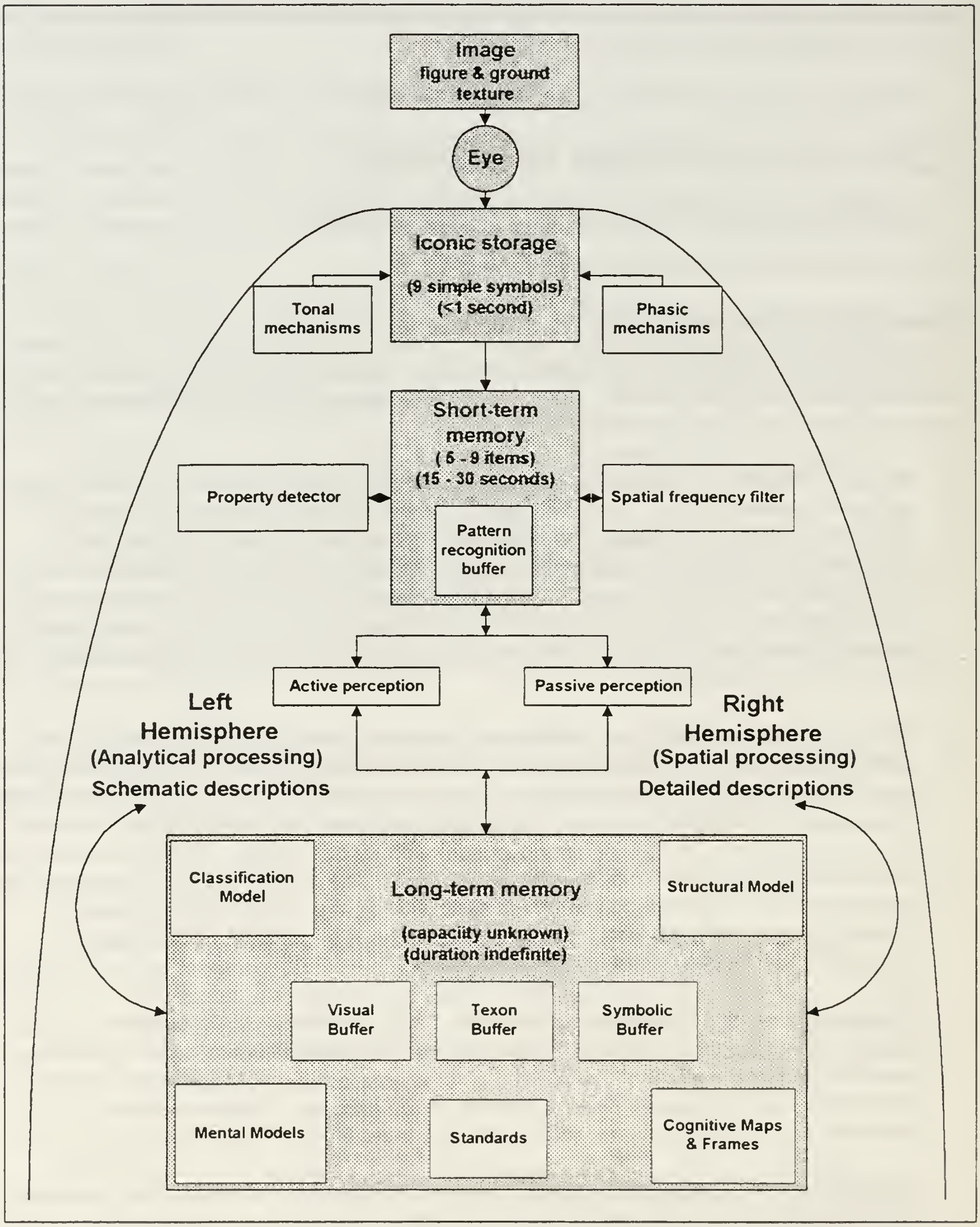

Figure 2. Components of Dudkin's 1985 Model of the Visual Process.

6. **Cerebral hemisphere asymmetry**. The right cerebral hemisphere is better adapted for spatial processing and the left for linguistic and analytical problems Information presented to the left or right visual field enters into the right or left hemisphere, respectively. The two hemispheres may

have different functions during image storage and recognition. The left serves for schematic, generalized (invariant) descriptions, while the right is better for specific, detailed description of objects. Thus the left hemisphere may use the *classification* model of information storage, while the right hemisphere uses the *structural* model.

7. **Cognitive maps**. Items that are stored in long-term memory include cognitive maps or *frames*, that is, generalized idealized diagrams based on the processes of perception, collection, and/or synthesis of sensory information. A given diagram is specific to what is perceived and can be modified by experience. It accepts information, and changes as a result of this information. It guides motions and investigative activities that enable access to new information which in turn further alters the diagram. The cognitive map takes an active, organizing role in the collection of sensory information and in comparing it with what is stored. Context is very important in this process.

8. **Mental models**. Pattern recognition is possible only because mental models of images (cognitive information structures), formed during the learning process, are stored in long-term memory. Impressions obtained during perception of an image and during reproduction of that image from memory appear to be identical. Time of reaction to a specific property of a mental image depends on the image's size and complexity. Small size and great complexity both increase reaction time.

9. **Mental models stored as images**. It has been proposed that images are stored in *visual form* in a visual buffer, including spatial metrics and all properties inherent in the actual visual object. The organization of the visual buffer is inborn and is limited in size. Information is represented in the buffer by selective activation of local regions of coordinate space in the cortex. The process of image formation includes: (1) generation of mental images, (2) checking conformity of the mental image against what is required, (3) transformations of the image (shifting, turning, scale changes), and (4) spontaneous evocation of information from long-term memory. It is assumed that the cortex-level buffer includes a "screen" on which long-term memory information (fragments of objects or scenes) is called up. A seemingly latent image of the mental image is formed on this screen.

10. **Mental models stored as symbols**. It also has been proposed that images are stored in *abstract symbolic form*, without spatial metrics. Images then are reproduced by processing and analysis of symbolic information, that is, information stored in a constant abstract symbolic format. Stored information is understood by means of grammar analysis procedures. The model is based on lists of "statements" (symbols), each of which is given a specific name.

11. **Retrieval from storage**. Indicators, primary elements, and prototypes are selected for retrieval from memory based on the *context* in which the object is embedded in storage. This simplifies the search in long-term memory, since it is possible to restrict the number of images that correspond to a given object.

Dudkin's major points related to the *perception and recognition process* can be summarized as follows.

1. ***Property detector* model**. The property detector model proposes that neurons or other inherent mechanisms automatically detect and distinguish various image properties during human visual analysis. Included automatic processes are those responsible for (1) perception and measurement of brightness and color, (2) perception of outline and gradient, (3) perception and measurement of

curvature, (4) perception of angles, intersections, and breaks in lines and boundaries, (5) perception of spots, (6) perception of texture (surface structure) and measurement of the texture gradient, and (7) detection and measurement of speed and acceleration. Other processes also may automatically distinguish (8) lines of different orientations, (9) location in the field of vision, and (10) object distance.

2. ***Spatial frequency filter* model**. The spatial frequency filter model of vision proposes that the distribution of illumination can be represented as a function based on superimpositions of harmonic components whose amplitudes are the Fourier coefficients of the function's spectrum. The spectrum is discrete for periodic functions and continuous for aperiodic functions. The visual system may include a set of independent channels, each with bandwidth between 0.5 and 1 octave and each tuned to a narrow spatial frequency range. The output from each channel is a threshold device for detection of a specific signal. Low-frequency components of the spectrum probably describe the object itself and high-frequency components describe the elements of the object. Very high spatial frequency components are necessary for pattern recognition. Filters possibly assist with several functions during the preattentional first stage of processing: (1) extract the figure from the background, (2) segment the images, (3) distinguish boundaries, and (4) separate out textural regions with different spatial frequency spectra.

3. **Combined models**. Evidence does not contradict either the visual model based on detectors or that based on spatial frequency filters. Most likely they complement each other. Apparently visual neuronal structures do not rely on one single, universal method for processing and inscribing sensory information.

4. ***Controlling* model of vision**. The controlling model, also called the *model of active perception*, proposes that controlling processes segment scenes to obtain optimum descriptions and to identify indicators needed for pattern recognition. Controlling processes can be restructured or readjusted through training and experience, resulting in formation of new mechanisms or revision of already available mechanisms. This model postulates the existence of two stages in perception: a preattention stage of parallel processing and an attentional stage of series processing. During the *preattention stage,* automatic mechanisms separate out the environment's specific properties and characteristics via a set of neuronal structures, either inherent or formed during the process of learning. These automatic mechanisms include both detectors of properties and spatial frequency filters. The *attentional stage* utilizes controlling processes that manipulate information flow to and from memory. These processes include coding, making decisions, and searching short-term and long-term memory. Sensory information is processed and memorized better in one situation than in another as a function of the control setting (adjustment) of the perceiver. Attention appears to be selective, so that information from the environment is received piecewise in a specific sequence.

5. ***Filtering* model of vision**. The filtering model or *model of passive perception* reduces the function of attention to filtering of some image properties and the blocking or considerable weakening of others. That is, neurons serve as selective filters that pass necessary signals on through and block the rest. Information is passed through various filters where complex conversions are carried out. Analyzing elements assign values to the determining indicators. At the highest level of perception, information about the indicators is compared with that previously accumulated in memory and conscious understanding of the synthesized image emerges.

6. **Standards**. Recognition perhaps is accomplished by comparing images with standards held in long-term memory. These standards are used for automatically distinguishing properties such as

object orientation, boundaries and outlines, line lengths, the position of an object in the field of vision, and its texture. Standards can be both inherent and learned.

7. **Figure and ground**. Objects are perceived as figures arranged against a background. Environmental and personal characteristics determine which objects are seen as figures and which as background. The visual mechanism separates figure from background by segmenting the visual scene. Boundaries of the objects that emerge as figures are detected, and fields with uniform distributions are separated from those that are textured.

8. **Texture**. Regions in an image that are perceived as uniform can be considered textures. Texture recognition probably occurs in the preattentive phase of vision as parallel automatic processing of images is taking place. One- or two-dimensional textures can have statistical connections of elements in two-dimensional space. *First-order statistics* characterize the distribution of combined probabilities that a point, randomly located on the texture, falls on the black (or white) color. *Second-order statistics* characterize the distribution of combined probabilities that both ends of a line will fall on the black (or white) color. Naturally-occurring textures can be described using these statistics. If figure and background can be characterized using first-order statistics, humans can easily and rapidly distinguish the figure (160 milliseconds). If second-order statistics must be used, the figure usually cannot be separated from the background even though the same value in the distribution is used as for first-order statistical descriptions.

9. **Texons**. Classes of textures called *texons* appear to be basic elements used by the visual system for automatically distinguishing textures. These structures may recognize quasi-linearity; closure; angle; connectivity; grain size; color; elongated "droplets" of specific orientation, width and length; and density of micropattern. An increase in the number of texons increases the probability of perception. Presence of texons in the figure improves perception more than their presence in the background.

This page intentionally left blank.

Vision and Thought

Introduction

Vision and Thought was published by V.D. Glezer in 1985. As its abstract notes, the book discusses the results of neurophysiological, behavioral, and model research into vision — which is defined as "objective thought." As in his 1961 text *Information and Vision*, Glezer emphasizes the application of Information Theory to human perception of objects, but now adds tenets from pattern recognition research along with the artificial intelligence domain. This book is much more physiological in emphasis than the earlier text, going into great detail concerning the visual cortex and its components, and how each component appears to process visually-acquired information.

Vision, equated by Glezer with thought, is considered to serve as a model of the world, stored in the brain in an orderly form that make access easy. Sensory images do not exist alone; they are separate from but closely linked with those motor actions that execute them (e.g., eye movements). The sensory image of the world arises from visual training. The whole sensory system participates in recognition of a given stimulus and creation of its patterns, using experience accumulated earlier and stored in memory.

Glezer claims that interactions between vision and thought result in three propositions: (1) a visual pattern is the result of sensory training, (2) this sensory training is not necessarily associated with behavioral actions, and (3) visual patterns are formed in specialized visual compartments of the brain. The book deals primarily with the third proposition, taking a neurophysiology approach. The author is interested in the structure of that part of the brain which creates the individual's model of the world.

Visual Cortex Receptor Fields and Modules

. Glezer cites two competing hypotheses regarding the nature of mental models of the visual world. The *detector hypothesis* proposes the existence of operators that distinguish the most frequently encountered elements of images: lines, corners, junctions, etc. This theory is widely accepted.

The *space-frequency hypothesis* suggests that a signal is encoded not at one point taken individually but is distributed with respect to its surroundings (piecewise or local description) or with respect to the entire field of view (global). According to this second hypothesis, cortical receptor fields describe images by expanding them in accordance with some system of basic functions (e.g., a trigonometric Fourier series). Glezer hypothesizes that a group of cortical neurons lying close together, with receptor fields projected onto one segment of the field of view and tuned to various orientations and spatial frequencies, carries out a Fourier description of this segment. He refers to such a group as a *module*, and cites numerous examples of research that supports such piecewise Fourier transforms in the visual system.

The Prestriate Cortex

The visual system performs two main tasks: (1) discrimination (classification) of observed objects and (2) description of spatial relationships among objects and object components. These processes are assumed to occur in various areas of the prestriate cortex. Based on laboratory data, Glezer proposes that the receptor fields of this portion of the brain carry out discrimination and description of subpatterns (areas with a common texture) that make up an object. A grating of modules, made up of complex non-linear fields, is superimposed on the visual field. The visual field then will be redescribed at the level of the modules and a new neuronal pattern will arise: in all modules lying within the limits of one texture, there will be identical patterns of excited neurons.

The Inferior Temporal Cortex and Posterior Parietal Cortex

Glezer proposes that classification of objects occurs in the inferior temporal cortex region of the brain and that description of spatial relationships is the responsibility of the posterior parietal cortex. That is, the former carries out invariant pattern recognition and the latter concretizes and describes the pattern completely.

The process of pattern recognition includes two stages: describing the image using a set of simple inborn signs, then making a decision regarding the pattern based on more complex signs generated from simpler ones in the process of visual training. Glezer suggests that simple signs are distinguished in the inferior temporal cortex as a result of image expansion by two-dimensional space-frequency channels. Very economical space-frequency descriptions will suffice to identify a pattern. It is beneficial to have such economical descriptions for storage and, more important, for retrieval from memory and comparison with incoming information — as is needed for the decision-making step (hypothesized to be carried out via a decision tree mechanism).

Glezer discusses the differing roles of the left and right cerebral hemispheres in pattern recognition. He considers that the mechanism of an invariant description of object shape, even with changes in size, is localized in the left hemisphere. The right hemisphere does not appear to possess invariant properties but instead remembers a specific figure of a specific size.

The posterior parietal cortex area, which stores models of intra- and extra-personal space, appears to be responsible for both selective attention (associated with eye movements) and description of spatial relationships. Glezer uses Minsky's concept of *frames* to explain the mechanism. A frame is a structure of data that describes some stereotype situation. When the visual system encounters a new situation, it compares the situation with frames stored in memory to make the internal model match reality by changing or replacing its parts as needed. If available information is insufficient, frame cells are filled in using selective attention (considered to be local activation of operators using feedback loops) to measure spatial properties.

Four Levels of the Visual System

Glezer ends his book with a summary of four levels of the visual system. The first level is the description of an image by the set of receptor fields of the visual subcortex. Neurons of the retina and lateral geniculate body measure the integrated light energy and carry out preliminary processing, including discriminating the signal from noise, emphasizing its contours and areas of high spatial frequency, and decorrelating the image with respect to space and time.

The second level is carried out in receptor fields of the striate cortex via two-dimensional spatial frequency filters. The fields measure the distribution of energy, and sets of similar fields form modules, each of which gives a local spectral description of its own section of the visual field. Some modules carry out piecewise Fourier descriptions of the image, others distinguish the amplitude of the space-frequency components (but lose the phase) to calculate the local power spectrum of a given section of the image, while still others react to the boundaries between sections of the image with differing spectral compositions.

At the third level, neurons of the prestriate cortex perform a base description of the image, including measurement of the neuron excitations from the previous level. These neurons also describe texturally homogeneous areas of the image (i.e., sections of the field of view with identical piecewise power spectra). At this level, the visual space is segmented into individual figure elements.

The fourth level of the visual system includes the inferior temporal cortex and posterior parietal cortex areas. This level contains the "training neurons," with training occurring in the simplest recurrent method, made possible by the opponent organization of the preceding three levels. In the right hemisphere inferotemporal cortex images are described and stored by combining figure elements distinguished by the prestriate cortex modules. In the left hemisphere the image is recognized via discrimination of complex distinctive signs or shapes.

While the inferotemporal area recognizes patterns, the posteroparietal area simultaneously uses frames to describe relationships. Spatial relationships between patterns or sub-patterns are described, along with logical relationships between objects. Frame cells are filled using the mechanism of selective attention. As a joint operation of the two cortex areas, both a generalized figurative abstract description and a concrete one are provided. Each act of visual perception includes comparison of new information with the entire ordered model of the world stored in the visual brain and, simultaneously, augmentation and further development of the model. This is Glezer's rationale for viewing the act of visual perception as an act of objective thought.

Vision and Thought: Model Usefulness

My review of *Vision and Thought* suggests that Glezer's 1985 four-level model can be summarized as illustrated in Figure 3. Glezer's concepts and his descriptions of the neurophysiological visual process, based on some 30 years of work at the Pavlov Institute, are indeed impressive when his text is studied in detail.

Unfortunately, most of the details included in this book (while very interesting to read) have little direct application to a practical model of vision for predicting target acquisition performance. Even if a model of vision *per se* were desired (somewhat like the British Aerospace *Oracle* model of visual perception, based primarily on physiological data), this text does not include sufficient details to construct the needed algorithms. Only a very few parameters are described in mathematical terms, and even those equations are very general in nature. The four-level model of the visual system discussed above possibly could serve as a framework on which eventually to build a computer model, but the amount of experimental research needed for data collection would likely be prohibitive.

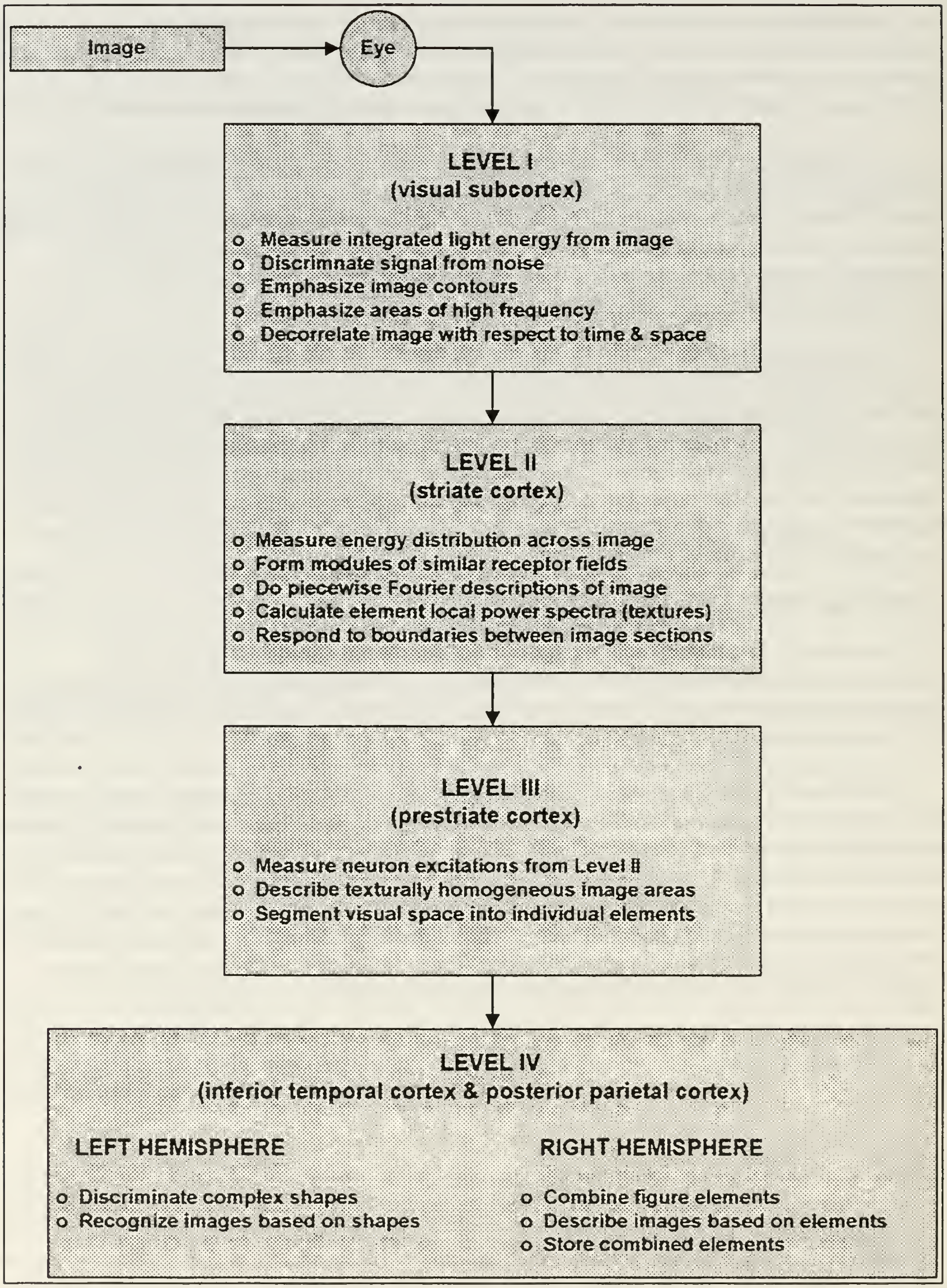
Image
Eye
LEVEL I
(visual subcortex)
o Measure integrated light energy from image
o Discrimnate signal from noise
o Emphasize image contours
o Emphasize areas of high frequency
o Decorrelate image with respect to time & space
LEVEL II
(striate cortex)
o Measure energy distribution across image
o Form modules of similar receptor fields
o Do piecewise Fourier descriptions of image
o Calculate element local power spectra (textures)
o Respond to boundaries between image sections
LEVEL III
(prestriate cortex)
o Measure neuron excitations from Level II
o Describe texturally homogeneous image areas
o Segment visual space into individual elements
LEVEL IV
(inferior temporal cortex & posterior parietal cortex)
LEFT HEMISPHERE
o Discriminate complex shapes
o Recognize images based on shapes
RIGHT HEMISPHERE
o Combine figure elements
o Describe images based on elements
o Store combined elements

Figure 3. Glezer's 1985 Four-level Model of the Visual Process.

Conclusions and Tentative Top-Level Model

The information included in the preceding sections has been reviewed to pick out factors that might be incorporated in a model of target acquisition based on Soviet visual perception concepts. At the present this is a very general and sketchy model. Some portions of it conflict with others, there is a considerable amount of overlap, and steps are missing. Some details perhaps could be filled in if additional pertinent information can be located. For now, a typical sequence of target acquisition events is used here to organize various model factors and components considered important by Glezer, Dudkin, and their associates.

Figure 4 provides a flowchart representation of the target acquisition process based on Soviet vision and cognitive concepts. The important factors to model for the *image* include an illumination function, exposure time, image apparent length and width, and target-to-background apparent contrast. Factors to model for the *eye* include accommodation, eye movements and fixations, reconfiguration of the receptive field, and temporal accumulation. An *iconic memory model* would include all scene elements that might be inspected by the observer during search. These elements are fed through a *filter* that consists of image and luminance segmentation, figure-ground separation, and texture regions. Moving into *short-term memory*, the factors presumed to influence detection, aimpoint, recognition, and identification are shown with each of these processes (and discussed below). *Long-term memory* effects of mental models, cognitive maps, standards, and indicators and elements complete the top-level model.

The following list may serve as a start towards later development of a comprehensive model; it also may include items that could be quantified and incorporated into existing U.S. military models. The various factors are separated into the categories of detection, aimpoint, recognition, and identification depending on the process they affect (with several factors included under more than one category). Once again, the model is incomplete and, for now, conceptual rather than quantitative.

Detection

1. **Simultaneous and subsequent contrast**. Perceived differences in luminosity of foreground objects due to background luminance and illuminance changes result in distortions of apparent contrast that can significantly affect whether a target object is perceived.

2. **Accommodation, Saccades, and Fixations**. The accommodation reflex is triggered and images are focused on the retina only when objects subtend at least 10 arc-min (although images can be detected even when accommodation is not complete). The eye may require two to four saccades before it is aimed at a point of interest. Micromovements of the eye are critical for adequate scene resolution. Between 15 and 25 characters can be scanned during one fixation. Asymptotic scan rate performance is approached when arrays are presented every 120 milliseconds, but the number of characters scanned does not increase beyond 240 milliseconds per fixation. Scan rates in excess of 100 characters per second can be achieved by most observers when new arrays are presented every 40 to 50 milliseconds for scanning..

3. **Model of illumination**. Distribution of illumination in a scene can be represented as a function based on superimpositions of harmonic components whose amplitudes are the Fourier coefficients of the function's spectrum. The spectrum is discrete for periodic functions and continuous for aperiodic functions. The visual system may include a set of independent channels, each with bandwidth between 0.5 and 1 octave and each tuned to a narrow spatial frequency range. The output from each channel is a threshold device for detection of a specific signal.

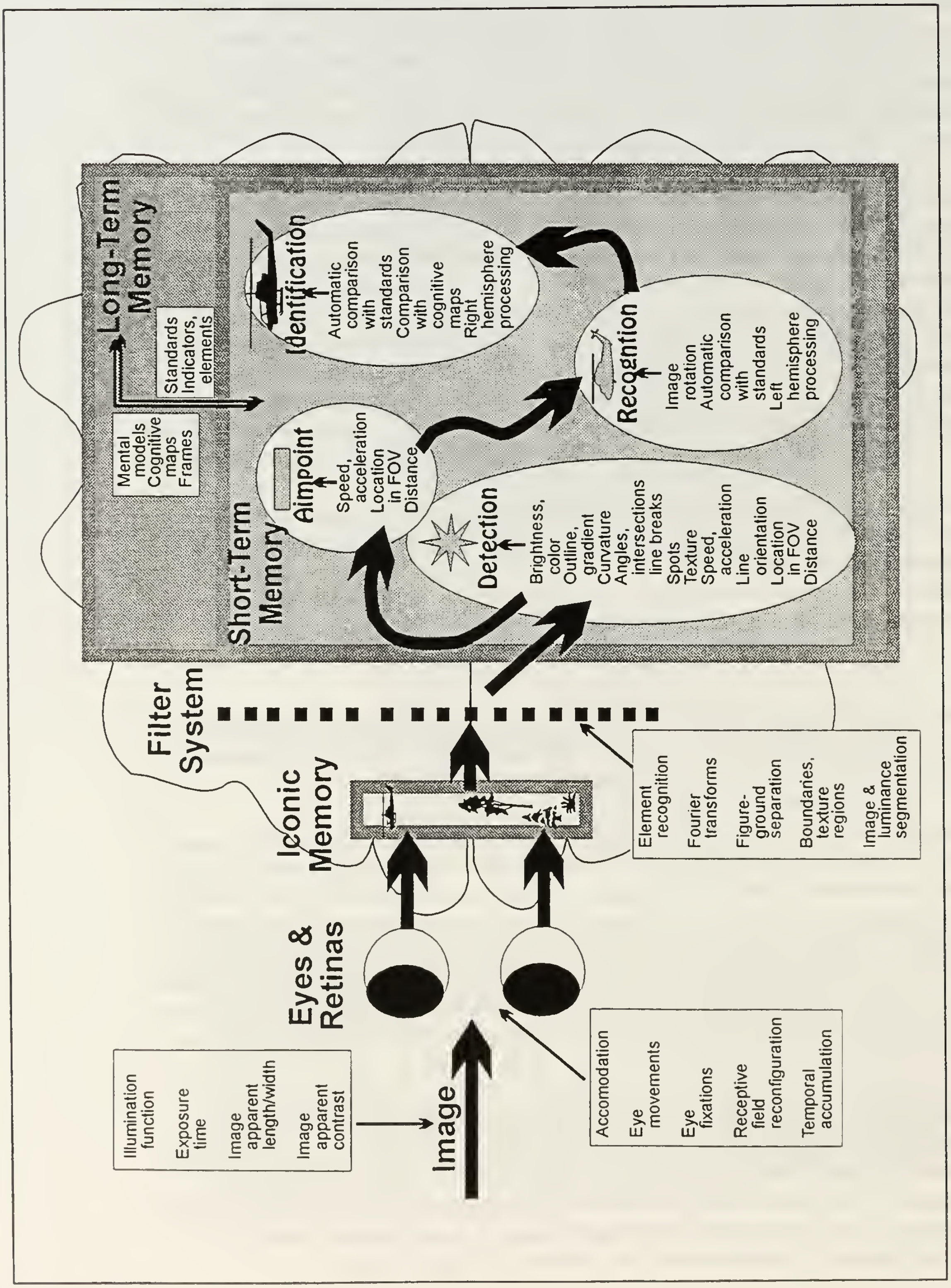

Figure 4. Top-Level Model of the Target Acquisition Process, Based on Soviet Visual Perception Concepts.

4. **Critical duration**. Receptive fields respond to a light stimulus when the critical minimum duration is exceeded, about 0.1 second at low light levels. The critical duration decreases somewhat as illumination increases. Stabilized images exposed on one spot on the retina for less than the critical duration have better resolution than non-stabilized images. When an image is exposed longer than the critical duration, natural eye movements increase resolution and a stabilized image has poorer resolution.

5. **Receptive field reconfiguration**. Foveal and peripheral receptive fields are reconfigured for various brightness levels. Rod receptive fields vary in size between 15 arc-min and 1 arc-deg. Foveal receptive fields vary between 5 arc-deg under scotopic conditions to the size of a single cone (0.4 arc-min) at average cloudy-day luminance levels. The size of the field determines the smallest object that can be discerned.

6. **Temporal Accumulation**. Receptive fields accumulate photons and sum them over the area of accumulation to result in a perceived light level that depends on stimulus size, luminance, and presentation time. With good contrast and reasonable light, about 0.08 second is needed for perception.

7. ***Controlling* model of vision**. Controlling processes segment scenes and identify indicators needed for pattern recognition. Controlling processes can be restructured or readjusted through training and experience. During the *preattention stage,* automatic mechanisms separate out the environment's specific properties and characteristics via a set of neuronal structures, either inherent or formed during the process of learning. These automatic mechanisms include both detectors of properties and spatial frequency filters. Attention appears to be selective, so that information from the environment is received piecewise in a specific sequence.

8. ***Filtering* model of vision**. The function of attention is to filter some image properties and block or weaken others. That is, neurons serve as selective filters that pass necessary signals on through and block the rest. Information is passed through various filters where complex conversions are carried out. Analyzing elements assign values to the determining indicators.

9. **Asymmetrical spatial summation**. In general, visibility of a solid rectangle of light increases when the shorter dimension of the stimulus is held constant and the longer dimension increases. There is a monotonic increase in visibility as length increases from 5 to 50 arc-min. The same monotonic increase occurs, but with shallower slope, when the width is 10 arc-min. When the width is 20 arc-min, visibility increases as length increases up to 40 arc-min, then decreases slightly as length increases further to 50 arc-min. When the width is wider than 20 arc-min, a similar pattern is observed. When the length is held constant and the width is increased, only between 0 and 5 arc-min does visibility increase. Beyond 5 arc-min there is little systematic relationship between visibility and width increase. The lack of a generally increasing relationship means that visibility cannot be a function either of total perimeter or total area *per se*.

10. **Figure and ground**. Objects are perceived as figures arranged against a background. Environmental and personal characteristics determine which objects are seen as figures and which as background. The visual mechanism separates figure from background by segmenting the visual scene. Boundaries of the objects that emerge as figures are detected, and fields with uniform distributions are separated from those that are textured.

11. **Texture**. Regions in an image that are perceived as uniform can be considered textures. Texture detection probably occurs in the preattentive phase of vision as parallel automatic processing of images is taking place. One- or two-dimensional textures can have statistical connections of elements in two-dimensional space. *First-order statistics* characterize the distribution of combined probabilities that a point, randomly located on the texture, falls on the black (or white) color. *Second-order statistics* characterize the distribution of combined probabilities that both ends of a line will fall on the black (or white) color. Naturally-occurring textures can be described using these statistics. If figure and background can be characterized using first-order statistics, humans can easily and rapidly distinguish the figure (160 milliseconds). If second-order statistics must be used, the figure usually cannot be separated from the background even though the same value in the distribution is used as for first-order statistical descriptions.

12. **Texons**. Classes of textures called *texons* appear to be basic elements used by the visual system for automatically distinguishing textures. These structures may detect quasi-linearity; closure; angle; connectivity; grain size; color; elongated "droplets" of specific orientation, width and length; and density of micropattern. An increase in the number of texons increases the probability of perception. Presence of texons in the figure improves perception more than their presence in the background.

13. **Discrete image elements**. Images are transmitted through the visual system (retina to optic centers of the brain) as a finite collection of discrete *elements*. A continuous two-dimensional image can be considered a distribution of various magnitudes of luminance on a surface, discretized by the visual system into a finite number of elements.

14. **Discrete luminance levels**. Light intensities are transmitted through the visual system (retina to optic centers of the brain) as a finite number of discrete *luminance levels*. The level of luminosity of the image is perceived at each discrete point that defines the object's length (or height). For low-light-level photopic vision, the number of luminous gradations is about nine. Discernible gradations of brightness can be considered *symbols*, and the collection of symbols is referred to as an *alphabet*.

15. **Visual detector model**. Neurons or other inherent mechanisms automatically detect and distinguish various image properties. Included automatic processes are those responsible for (1) perception and measurement of brightness and color, (2) perception of outline and gradient, (3) perception and measurement of curvature, (4) perception of angles, intersections, and breaks in lines and boundaries, (5) perception of spots, (6) perception of texture (surface structure) and measurement of the texture gradient, and (7) detection and measurement of speed and acceleration. Others also may automatically distinguish (8) lines of different orientations, (9) location in the field of vision, and (10) object distance.

16. **Visual analyzer throughput capacity**. At low illumination levels, throughput capacity grows linearly with logarithmic increase of illumination. An illumination increase of 2 results in throughput capacity growth of about 10 bits per second. At higher levels of illumination, throughput capacity becomes constant, but capacity depends on the task.

17. **Channels and bandwidths**. The visual system may include a set of independent channels, each with bandwidth between 0.5 and 1 octave and each tuned to a narrow spatial frequency range. The output from each channel is a threshold device for detection of a specific signal. Low-frequency components of the spectrum probably describe the object itself and high-frequency components

describe the elements of the object. Filters possibly assist with several functions during the preattentional first stage of processing: (1) extract the figure from the background, (2) segment the images, (3) distinguish boundaries, and (4) separate out textural regions with different spatial frequency spectra.

18. **Memory structures**. The memory system used for detection most likely is the *iconic* system (duration < 1 second, capacity 9 simple symbols). Information passes from the iconic memory buffer to short-term memory. Control processes coordinate information flow throughout the overall system.

19. **Time to detect**. An image must be exposed for at least 36 ± 0.8 milliseconds for perception to occur (but see Item 6 above).

Aimpoint

1. **Memory structures**. The memory system used for the aimpoint process most likely is the *iconic* system (duration < 1 second, capacity 9 simple symbols). Information passes from the iconic memory buffer to short-term memory. Control processes coordinate information flow throughout the overall system.

2. **Time of inertia**. Perception of the image is maintained in the optic system for 0.012 to 0.2 second after the stimulus disappears for the fovea, and from 0.1 to 0.32 second for peripheral vision. Time of inertia increases as stimulus intensity increases.

Recognition

1. **Memory structures**. All three types of memory systems probably are used in conjunction with recognition: *iconic* (duration < 1 second, capacity 9 simple symbols), *short-term* (duration 15 to 30 seconds, capacity 5 to 9 items), and *long-term* (duration indefinite, capacity unknown). Information passes from the iconic memory buffer to short-term memory, which includes the working buffer used for pattern recognition. Long-term memory passes appropriate information to short-term memory for correlation with what is perceived.

2. **Temporal Accumulation**. Receptive fields accumulate photons and sum them over the area of accumulation to result in a perceived light level that depends on stimulus size, luminance, and presentation time. With good contrast and reasonable light about 0.25 second is needed for recognizing an object's general outline.

3. **Visual analyzer throughput capacity**. Individual letters can be recognized at rates of about 70 bits per second. Image recognition can occur at throughput rates of 50 to 70 bits per second. Reading speed is measured at 30 to 40 bits per second.

4. **Number of objects perceived**. After coding and transmission through the optic system, the observer perceives only about 7 ± 2 objects simultaneously over a 0.1-second period of time. Information is stored in short-term memory for 0.27 second, then disappears unless the rehearsal or repetition process is used to prolong its retention.

5. **Channels and bandwidths**. The visual system may include a set of independent channels, each with bandwidth between 0.5 and 1 octave and each tuned to a narrow spatial frequency range. The output from each channel is a threshold device for detection of a specific signal. Low-frequency

components of the spectrum probably describe the object itself and high-frequency components describe the elements of the object.

6. *Controlling* **model of vision**. Controlling processes identify indicators needed for pattern recognition. Controlling processes can be restructured or readjusted through training and experience. The *attentional stage* utilizes controlling processes that manipulate information flow to and from memory. These processes include coding, making decisions, and searching short-term and long-term memory. Sensory information is processed and memorized better in one situation than in another as a function of the control setting (adjustment) of the perceiver. Attention appears to be selective, so that information from the environment is received piecewise in a specific sequence.

7. *Filtering* **model of vision**. The function of attention is to filter some image properties and block or weaken others. That is, neurons serve as selective filters that pass necessary signals on through and block the rest. Information is passed through various filters where complex conversions are carried out. Analyzing elements assign values to the determining indicators. At the highest level of perception, information about the indicators is compared with that previously accumulated in memory and conscious understanding of the synthesized image emerges.

8. **Short-term storage**. Retention of information in short-term memory depends substantially on its spatial frequency content. High-frequency information (contrast boundaries, outlines, fine details, some fine-grained textures) is retained up to 10 times longer than low frequency information (large regions of uniform brightness, smooth illumination gradients, low-frequency textures).

9. **Configuration-sensitive receptive fields**. Certain receptive fields automatically discern specific simple configurations (shapes) that make up images and codify each as an entire simple configuration The alphabet of shapes for this process is partly inherent and partly acquired on an individual basis through training and experience.

10. **Standards**. Recognition perhaps is accomplished by comparing images with *standards* held in long-term memory. These standards are used for automatically distinguishing properties such as object orientation, boundaries and outlines, line lengths, the position of an object in the field of vision, and its texture. Standards can be both inherent and learned.

11. **Outlines**. A shaded image can be replaced by an outline of the same object, with areas of transition from black to white (or between shades of gray) marked by borders. The resultant outline will carry essentially the same amount of information as the shaded image but can be transmitted much more efficiently. Emphasized borders enhance apparent contrast of an object, especially when exposure time is longer than the critical duration and the image is not stabilized to counteract normal eye movements. Borders that change direction sharply or that outline fine details are especially important for drawing visual attention and for discrimination.

12. **Classification model**. Each image may be stored in long-term memory in the form of a multidimensional vector of indicators. Pattern recognition consists of picking out the appropriate regions of the space of all stored indicators, with the aid of specific separating functions. Long-term memory possibly contains various systems or sets of indicators. Which is selected is determined by the type of perceived sensory information and by the method of recognition.

13. **Structural model**. Images may be stored in long-term memory in the form of symbolic structural descriptions of objects — idealized schema that include sets of primary elements and their grammars and that assign rules of grammatical selection during analysis and pattern recognition. The specific sets of primary elements depend on the type of perceived sensory data.

14. **Cerebral hemisphere asymmetry**. The two hemispheres may have different functions during image recognition. The left serves for schematic, generalized (invariant) descriptions, while the right is better for specific, detailed description of objects. Thus the left hemisphere may use the *classification* model of information storage, while the right hemisphere uses the *structural* model.

15. **Cognitive maps**. Generalized idealized diagrams (also called *frames*) guide motions and investigative activity that enable access to new information and allow the observer to compare perceived images with stored images. Context is very important in this process.

16. **Mental models**. Pattern recognition is possible only because mental models (cognitive information structures), formed during the learning process, are stored in long-term memory. Impressions obtained during perception of an image and during reproduction of that image from memory appear to be identical. Time of reaction to a specific property of a mental image depends on the image's size and complexity. Small size and great complexity both increase reaction time.

17. **Retrieval from storage**. Indicators, primary elements, and prototypes are selected for retrieval from memory based on the *context* in which the object is embedded in storage. During pattern recognition, this simplifies the search in long-term memory, since it is possible to restrict the number of images that correspond to a given object.

18. **Time to Recognize**. An image must be exposed for at least 36 ± 0.8 milliseconds for perception to occur. The entire recognition process takes between 226 ± 15.6 milliseconds and 414 ± 17.0 milliseconds (but see Item 2 above).

19. **Rotated Images**. When an image is rotated, accuracy of recognition is not affected, up to 10 to 15 degrees from the learned orientation. With larger changes in orientation, there is an approximately linear dependence between the angle of rotation and probability of correct recognition. If orientation is changed by as much as 30 to 60 degrees, results vary for various figures and various observers. Figures then may not be recognized at all with presentation times of 100 to 150 milliseconds, but can be recognized given long presentation times. If figures are rotated no more than 15 degrees, correct "same-different" responses generally are given in 300 to 450 milliseconds. For large differences in orientation, response times grow significantly, so that rotations of 120 degrees require between about 500 and 650 milliseconds for figure recognition.

Identification

1. **Memory structures**. Both *short-term* (duration 15 to 30 seconds, capacity 5 to 9 items), and *long-term* (duration indefinite, capacity unknown) memory probably are used for identification. Information passes from the short-term store to long-term memory. In turn, long-term memory passes appropriate information to short-term memory for correlation with what is perceived.

2. **Temporal Accumulation**. Receptive fields accumulate photons and sum them over the area of accumulation to result in a perceived light level that depends on stimulus size, luminance, and

presentation time. With good contrast and reasonable light about 0.6 second is needed for identifying an object.

3. **Short-term storage**. Retention of information in short-term memory depends substantially on its spatial frequency content. High-frequency information (contrast boundaries, outlines, fine details, some fine-grained textures) is retained up to 10 times longer than low frequency information (large regions of uniform brightness, smooth illumination gradients, low-frequency textures).

4. **Long-term storage**. Information about objects, including their characteristics, qualities, states, and interdependencies, is entered into storage along with (1) what preceded this impression, (2) what accompanied it, (3) what followed it, and (4) with what it is similar, in whole or in part. These "keys" or "handles" assist in information retrieval.

5. **Classification model**. Each image may be stored in long-term memory in the form of a multidimensional vector of indicators. Pattern identification consists of picking out the appropriate regions of the space of all stored indicators, with the aid of specific separating functions. Long-term memory possibly contains various systems or sets of indicators. Which set is selected is determined by the type of perceived sensory information and by the method of identification.

6. **Structural model**. Images may be stored in long-term memory in the form of symbolic structural descriptions of objects — idealized schema that include sets of primary elements and their grammars and that assign rules of grammatical selection during analysis. The specific sets of primary elements depend on the type of perceived sensory data.

7. ***Filtering* model of vision**. The function of attention is to filter some image properties and to block or weaken others. That is, neurons serve as selective filters that pass necessary signals on through and block the rest. Information is passed through various filters where complex conversions are carried out. Analyzing elements assign values to the determining indicators. At the highest level of perception, information about the indicators is compared with that previously accumulated in memory and conscious understanding of the synthesized image emerges.

8. **Cerebral hemisphere asymmetry**. The right cerebral hemisphere is better adapted for spatial processing and the left for linguistic and analytical problems Information presented to the left or right visual field enters into the right or left hemisphere, respectively. The two hemispheres may have different functions during image identification. The left serves for schematic, generalized (invariant) descriptions, while the right is better for specific, detailed description of objects. Thus the left hemisphere may use the *classification* model of information storage, while the right hemisphere uses the *structural* model.

9. **Cognitive maps**. Items that are stored in long-term memory include *cognitive maps* (also called *frames*), that is, generalized idealized diagrams based on the processes of perception, collection, and/or synthesis of sensory information. The diagram is specific to what is perceived and can be modified by experience. It accepts information, and changes as a result of this information It guides motions and investigative activities that enable access to new information which in turn further alters the diagram. The cognitive map takes an active, organizing role in the collection of sensory information and in comparing it with what is stored. Context is very important in this process.

10. **Mental models of images**. Identification is possible only because cognitive information structures (mental models), formed during the learning process, are stored in long-term memory. Impressions obtained during perception of an image and during reproduction of that image from memory appear to be identical. Time of reaction to a specific property of a mental image depends on the image's size and complexity. Small size and great complexity both increase reaction time.

11. **Retrieval from storage**. Indicators, primary elements, and prototypes are selected for retrieval from memory based on the *context* in which the object is embedded in storage. This simplifies the search in long-term memory, since it is possible to restrict the number of images that correspond to a given object.

This page intentionally left blank.

Initial Distribution List

1. Research Office (Code 08) 1
   Naval Postgraduate School
   Monterey, CA 93943-5000

2. Dudley Knox Library (code 52) 2
   Naval Postgraduate School
   Monterey, CA 93943-5000

3. Defense Technical Information Center 2
   8725 John J. Kingman Rd., STE 0944
   Ft. Belvoir, VA 22060-6218

4. CAPT (Sel.) Frank Petho (Code OR/Pe) 1
   Naval Postgraduate School
   Monterey, CA 93943-5000

5. Department of Operations Research 1
   Editorial Assistant (Code OR/Bi)
   Naval Postgraduate School
   Monterey, CA 93943-5000

6. Prof. Judith H. Lind (Code OR/Li) 5
   Naval Postgraduate School
   Monterey, CA 93943-5000

7. US Army TRADOC Analysis Command 79
   Attn: ATRC-WBD, Bob Bennett
   White Sands Missile Range, NM 88002-5502

8. LTC Ralph Wood 1
   TRAC Monterey
   Naval Postgraduate School
   Monterey, CA 93943-5000

9. MAJ Bill Murphy 1
   TRAC Monterey
   Naval Postgraduate School
   Monterey, CA 93943-5000

10. Prof. Alan Washburn (Code OR/Ws) 1
    Naval Postgraduate School
    Monterey, CA 93943-5000

11. LTC Mark Youngren (Code OR/Ym) 1
Naval Postgraduate School
Monterey, CA 93943-5000

12. Prof. James G. Taylor (Code OR/Tw) 1
Naval Postgraduate School
Monterey, CA 93943-5000

13. Prof. Morris Driels (Code ME/Dr) 1
Naval Postgraduate School
Monterey, CA 93943-5000

14. LT. Kip Krebs (Code 034) 1
Naval Postgraduate School
Monterey, CA 93943-5000

15. Capt. Mat Sampson (Code 30) 1
Naval Postgraduate School
Monterey, CA 93943-5000

16. Mallory Boyd 1
Code 455530D
Naval Air Warfare Center Weapons Division
China Lake, CA 93555

17. Dale Robison 1
Code 455530D
Naval Air Warfare Center Weapons Division
China Lake, CA 93555

18. Dr. Robert Osgood 1
Code 455530D
Naval Air Warfare Center Weapons Division
China Lake, CA 93555

19. Jan Stiff 1
Code 455530D
Naval Air Warfare Center Weapons Division
China Lake, CA 93555

20. Army Research Laboratory 1
Attn: AMSRL-HR-SD (J. Swoboda)
Aberdeen Proving Ground, MD 21005

21. Ronald A. Erickson 1
ASI Systems International
2835 Loraine Dr.
Missoula, MT 59803